THE ANNALS
of
MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930-1938
EDITED BY S. S. WILKS, 1938-1949

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XXI

1950

THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Business Administration Building, University of Michigan, Ann Arbor, Michigan
C. H Fischer, Secretary-Treasurer
This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but *not* for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae. Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Price: $10.00 per year inside the Western Hemisphere; $5.00 elsewhere. Single issues $3.00. Back numbers are available at $10.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

Insuran

THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, 1930–1938
EDITED BY S. S. WILKS, 1938–1949

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. 21, No. 1 — March, 1950

THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Business Administration Building, University of Michigan, Ann Arbor, Michigan
C. H Fischer, Secretary-Treasurer
This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but *not* for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.
Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Price: $10.00 per year inside the Western Hemisphere; $5.00 elsewhere. Single issues $3.00. Back numbers are available at $10.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

te
te
p
m
w
p
e
e
s
ti

a
(
(
a
i
fa
x
a
n
(
t

h
t
U
v
u
i
I
-

c

SOME PRINCIPLES OF THE THEORY OF TESTING HYPOTHESES[1]

By E. L. Lehmann

University of California, Berkeley

Introduction:

1. The likelihood ratio principle. The development of a theory of hypothesis testing (as contrasted with the consideration of particular cases), may be said to have begun with the 1928 paper of Neyman and Pearson [16]. For in this paper the fundamental fact is pointed out that in selecting a suitable test one must take into account not only the hypothesis but also the alternatives against which the hypothesis is to be tested, and on this basis the likelihood ratio principle is proposed as a generally applicable criterion. This principle has proved extremely successful; nearly all tests now in use for testing parametric hypotheses are likelihood ratio tests, (for an extension to the non-parametric case see [33]), and many of them have been shown to possess various optimum properties.

At least in the parametric case the likelihood ratio test has a number of desirable properties. Among these we mention:

(i) Frequently it is easy to apply and leads to a definite and reasonable test.

(ii) If the sample size is large, and if certain regularity conditions are satisfied an approximate solution can be given for the distribution problems that arise in the determination of size and power of the test (Wilks [32], Wald [25]). In fact, if the likelihood ratio is denoted by λ, $-2 \log \lambda$ approximately has a central χ^2-distribution under the hypothesis, a non-central χ^2-distribution under the alternatives. The number of degrees of freedom in these distributions equal the number of constraints imposed by the hypothesis.

(iii) As was shown by Wald [25], under certain restrictions the likelihood ratio test possesses various pleasant large sample properties.

In view of this, one may feel that the likelihood ratio principle, although perhaps not always leading to the optimum test, is completely satisfactory, and that a more systematic study of the problem of test selection is not necessary. Unfortunately, against the pleasant properties just mentioned there stands a very unpleasant one. Cases exist, in which the likelihood ratio test is not only unsatisfactory but worse than useless, and hence the likelihood ratio principle is not reliable. Examples of this kind were constructed independently by H. Rubin and C. Stein; the following is Stein's example.

---

[1] Parts of this paper were presented in an invited address at the meeting of the Institute of Mathematical Statistics on Dec. 30, 1948, in Cleveland, Ohio.

Let X be a random variable capable of taking on the values 0, ± 1, ± 2 with probabilities as indicated:

	-2	2	-1	1	0
Hypothesis H:	$\frac{\alpha}{2}$	$\frac{\alpha}{2}$	$\frac{1}{2} - \alpha$	$\frac{1}{2} - \alpha$	α
Alternatives:	pC	$(1 - p)C$	$\frac{1 - C}{1 - \alpha}\left(\frac{1}{2} - \alpha\right)$	$\frac{1 - C}{1 - \alpha}\left(\frac{1}{2} - \alpha\right)$	$\alpha \frac{1 - C}{1 - \alpha}$

Here α, C are constants, $0 < \alpha \leqq \frac{1}{2}$, $\frac{\alpha}{2-\alpha} < C < \alpha$, and p ranges over the interval [0, 1].

It is desired to test the hypothesis H at significance level α. The likelihood ratio test rejects when $X = \pm 2$, and hence its power is C against each alternative. Since $C < \alpha$, this test is literally worse than useless, for a test with power α can be obtained without observing X at all, simply by the use of a table of random numbers. It is worth noting that the test, which rejects H when $X = 0$, has power $\alpha \dfrac{1 - C}{1 - \alpha} > \alpha$, so that a reasonable test of the hypothesis in question does exist.

The existence of such examples gives added importance to the problem of developing a systematic theory of hypothesis testing. It is the purpose of the present paper to give a brief survey of the work done on some aspects of such a theory and to indicate certain extensions and modifications of the existing theory. Some examples and applications will be considered. These will be restricted to parametric problems. For applications to testing non-parametric hypotheses see [12].

The results of sections 5 and 8 were obtained jointly by Gilbert Hunt and Charles Stein in 1945. They have not been published and were communicated to me by Professor Stein. I should like to express to him my gratitude for acquainting me with this material and for giving me permission to include it in this paper. I should also like to acknowledge my indebtedness to Professor Henry Scheffé who read the manuscript and made many helpful suggestions.

2. Formulation of the problem. The problem of testing a statistical hypothesis was formulated by Neyman and Pearson [18] as follows.

A random variable X is known to be distributed over a space $\mathfrak{X}$ according to some member of a family of probability distributions $\{P_\theta^X\}$, $\theta \in \Omega$. It will be assumed here that there is specified an additive class $\mathfrak{B}$ of sets in $\mathfrak{X}$, and that the probability distributions P_θ^X are probability measures defined over $\mathfrak{B}$. All sets or real valued functions mentioned in this paper will be assumed measurable $\mathfrak{B}$ unless otherwise stated. If $B \in \mathfrak{B}$, we shall write for the measure assigned to B by P_θ^X interchangeably $P_\theta^X(X \in B)$, $P_\theta^X(B)$, and if there is no possibility of confusion, $P_\theta(B)$. Throughout most of the paper it will be assumed that the probability measures P_θ^X are absolutely continuous with respect to a

given sigma finite measure μ defined over $\mathfrak{B}$, so that there exist non-negative functions f_θ such that

$$P_\theta(B) = \int_B f_\theta(x)\, d\mu(x). \tag{2.1}$$

Radon - Nikodym Thm

We shall then say that $f_\theta(x)$ is a generalized probability density *w.r.* to μ.

A statistical hypothesis H specifies a subset ω of Ω, and states that the distribution of X is some P_θ^X with $\theta \in \omega$. A test of H is any subset w of $\mathfrak{X}$, the convention being that H is rejected if the observed value x of X is in w, and that in the contrary case H is accepted. The selection of w is to be made as follows. A number α is given, $0 < \alpha < 1$, the level of significance, and w must be such that

$$P_\theta(w) = \alpha \text{ for all } \theta \in \omega. \tag{2.2}$$

Subject to this restriction it is desired to maximize $P_\theta(w)$ for θ in $\Omega - \omega$. The interpretation of these conditions is immediate. Since $P_\theta(w)$ is the probability of rejecting H computed under the assumption that P_θ^X is the distribution of X, equation (2.2) states that the probability of rejecting H is to be α (usually some small number such as .01 or .05) whenever H is true. Similarly the second condition expresses the fact that H is to be rejected with high probability when θ is in $\Omega - \omega$.

Naturally the second condition is not to be taken literally but rather as a loosely stated principle of choice. For in general there will exist a unique set w maximizing $P_{\theta_1}(w)$ for any given $\theta_1 \in \Omega - \omega$, but this w will change with θ_1. The condition has a clear meaning only in the case that the set $\Omega - \omega$ contains only a single point, and in a few special problems in which the same set w maximizes $P_\theta(w)$ for all $\theta \in \Omega - \omega$. In the general case there are available two main methods for making the condition precise. One may restrict consideration to some class of "nice" tests, so that within this class the maximization of $P_\theta(w)$ can be achieved uniformly for $\theta \in \Omega - \omega$. Alternatively, instead of asking that a local optimum property hold uniformly, one may look for a test whose power function possesses some optimum property in the large. Both of these approaches have an element of arbitrariness. In the first, the selection of a class of nice tests, in the second, the choice of an appropriate optimum property. Fortunately, in a number of important special cases, both methods, for various reasonable definitions, lead to the same test.

Before proceeding with this development, we shall modify the formulation of the problem slightly. First, as has been pointed out by many writers, it seems more natural to replace (2.2) by

$$P_\theta(w) \leqq \alpha \text{ for all } \theta \in \omega. \tag{2.3}$$

Secondly, we shall permit "randomized" tests (see [11, 29]), that is, instead of demanding that the statistician decide for each value of x whether to accept or to reject H, we shall allow the possibility that for certain x the decision be

reached by means of some chance device such as a table of random numbers. By a test of H we shall therefore mean a function ϕ from $\mathfrak{X}$ to the interval $[0, 1]$, with the convention that when x is the observed value of X some chance experiment with two possible outcomes R, $\bar{R}$ will be performed such that $P(R) = \phi(x)$, and that H will be rejected when the outcome is R and will otherwise be accepted. The case of a non-randomized test w clearly is obtained as a special case by taking for ϕ the characteristic function of the set w.

For a test ϕ the probability of rejection is given by

$$\int_{\mathfrak{X}} \phi(x)\, d\, P_\theta^X(x) = E_\theta \phi(X) \tag{2.4}$$

where E_θ denotes expectation computed with respect to the probability distribution P_θ^X. We therefore obtain the following formulation of the problem: To determine a test function ϕ $(0 \leqq \phi(x) \leqq 1)$ which maximizes $E_\theta\, \phi(X)$, the *power* of ϕ against the alternative θ, for θ in $\Omega - \omega$ subject to the condition

$$E_\theta \phi(X) \leqq \alpha \text{ for all } \theta \in \omega. \tag{2.5}$$

In this connection it is convenient to use the term *"level of significance"* for the preassigned number α, and to define the *size* of the test ϕ as

$$\sup_{\theta \in \omega} E_\theta \phi(X). \tag{2.6}$$

Except in the trivial case that there exists a test of size $< \alpha$ whose power is 1 against all alternatives, the size of any optimum test (in fact, of any admissible test) equals the level of significance.

3. Testing against a simple alternative. A complete solution of the problem formulated in the last section is available only in the case that ω and $\Omega - \omega$ each contains only a single point, that is, in the case that both the hypothesis and the alternative are simple. The solution is then given by the fundamental lemma of Neyman and Pearson [18], which we may state in the following slightly more complete form.

THEOREM 3.1. *Let*

$$P_\theta(A) = \int_A f_\theta(x)\, d\mu(x). \tag{3.1}$$

(a) *For testing the hypothesis H: $\theta = \theta_0$ against the alternative $\theta = \theta_1$ at level of significance α, there exists a number k and a test ϕ of size α such that*

$$\begin{aligned} \phi(x) &= 1 \quad \textit{when} \quad f_{\theta_1}(x) > k f_{\theta_0}(x), \\ \phi(x) &= 0 \quad \textit{when} \quad f_{\theta_1}(x) < k f_{\theta_0}(x). \end{aligned} \tag{3.2}$$

(b) *If $f_{\theta_0}(x)$ and $f_{\theta_1}(x)$ are $\neq 0$ for all x in $\mathfrak{X}$, then a test ϕ is most powerful for testing H against $\theta = \theta_1$ if and only if it satisfies* (3.2) *except possibly on a set of $\mu-$ measure* 0[2]. (Note that the number k of (3.2) is essentially unique).

[2] Throughout the paper we shall consider two tests as equal if they differ only on a set of μ-measure 0.

The second half of the theorem may be paraphrased by saying that under the conditions stated the most powerful test is uniquely determined by (3.2) except on the set on which

$$f_{\theta_1}(x) = k f_{\theta_0}(x). \tag{3.3}$$

On this set the value of ϕ may be assigned arbitrarily provided the resulting test has size α. If in particular the set on which (3.3) holds has measure 0, the most powerful test is unique.

It should be mentioned that (3.1) is no restriction since any two probability measures P_1, P_2 defined over a common additive class can be represented in this form with $\mu = P_1 + P_2$. If the assumption of (b) is not satisfied, the theorem is still true in essence but some trivial modifications are necessary.

No such complete solution is available for the problem of testing a composite hypothesis against a simple alternative. However, as was shown in [11], this problem may in many cases be reduced to the one just considered. Let the hypothesis state that θ is an element of ω, and consider the simple alternative $\theta = \theta_1$. Suppose that an additive class of sets has been defined on ω (in most of the applications ω is a subset of Euclidean space, and the additive class is formed by the Borel sets contained in ω). Then for any probability distribution λ over ω,

$$h_\lambda(x) = \int_\omega f_\theta(x)\, d\lambda(\theta) \tag{3.4}$$

is a probability density function with respect to μ.

Under certain conditions to be stated below, the most powerful test ϕ_λ for testing the simple hypothesis H_λ that X is distributed with probability density h_λ against the alternative f_{θ_1} is also most powerful for testing the original hypothesis H against the same alternative. This is essentially the Bayes approach developed by Wald for his general decision theory, and in fact, under the conditions which we shall state, λ is a least favorable distribution over ω in the following sense. Let β_λ be the power of ϕ_λ against f_{θ_1}, and for any distribution λ^* over ω denote by H_{λ^*}, ϕ_{λ^*}, β_{λ^*} the associated hypothesis, the most powerful test for testing it against f_{θ_1}, and the power of this test respectively. Then λ is said to be least favorable if for all λ^*

$$\beta_\lambda \leqq \beta_{\lambda^*}. \tag{3.5}$$

THEOREM 3.2. *Suppose there exists a probability distribution λ over ω such that the most powerful test ϕ_λ of size α for testing H_λ against f_{θ_1} is of size α also with respect to the original hypothesis H. Then*

(i) *ϕ_λ is most powerful for testing H against f_{θ_1};*

(ii) *λ is a least favorable distribution.*

Also, if ϕ_λ is the unique most powerful test for testing H_λ against f_{θ_1}, it is the unique most powerful test for testing H against f_{θ_1}.

These results are essentially contained in Wald's work (see for example theorem 4.8 of [26]).

There are many trivial applications of this theorem to finding most powerful tests of one-sided hypotheses concerning a single real-valued parameter, such as testing $H: p \leqq p_0$ against $p = p_1(p_0 < p_1)$ when X has a binomial distribution with parameter p. As is well known, it turns out in a number of these cases that the most powerful tests are in fact uniformly most powerful against the one-sided class of alternatives.

In [11] Theorem 3.2 was used to determine most powerful tests of certain hypotheses concerning normal distributions. As an example consider the case that $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and variance σ^2. Denote by H_1 and H_2 the hypotheses $\sigma = 1$ and $\xi = 0$ respectively, and let the alternative be: $\xi = \xi_1, \sigma^2 = \sigma_1^2$. Then the most powerful test of H_1 rejects if

$$(3.6) \qquad \begin{aligned} &\Sigma(x_i - \xi_1)^2 < k_1 \quad \text{when} \quad \sigma_1 < 1, \\ &\Sigma(x_i - \bar{x})^2 > c_1 \quad \text{when} \quad \sigma_1 > 1, \end{aligned}$$

and accepts otherwise. Here k_1 and c_1 depend only on the level of significance, that is, are independent of ξ_1, σ_1. If $\xi_1 > 0$, the most powerful test for testing H_2 rejects if

$$(3.7) \qquad \begin{aligned} &\Sigma(x_i - b)^2 \leqq k_2 b^2 \quad \text{when} \quad \alpha < \tfrac{1}{2}, \\ &\frac{\bar{x}}{\sqrt{\Sigma(x_i - \bar{x})^2}} \leqq c_2 \quad \text{when} \quad \alpha \geqq \tfrac{1}{2}, \end{aligned}$$

and accepts H_2 otherwise. Here k_2 and c_2 depend only on α, while b depends on ξ_1, σ_1 and α.

These results indicate that even when the class of alternatives is larger than in the above problems, some improvement over the standard tests may be possible provided good power is desired only against a narrow class of alternatives.

4. Sufficient statistics. Before treating the problem of composite alternatives, we shall consider an important simplification that can be obtained by making use of sufficient statistics. This notion was introduced by R. A. Fisher, and was further developed by J. Neyman [13] and in [2] and [10]. Consider any measurable partition of $\mathfrak{X}$. For any point x in $\mathfrak{X}$, let $t(x)$ be that set of the partition in which x lies. A set in the range of t is said to be measurable if the corresponding set of points x is an element of $\mathfrak{B}$. Denote the class of measurable t-sets by $\mathfrak{A}$. Then the *statistic* $T = t(X)$ is a random variable defined over $\mathfrak{A}$. Kolmogoroff has shown how for any $B \in \mathfrak{B}$ one can define the conditional probability $P(B \mid t)$ of B given $T = t$ uniquely up to a set of measure zero by the equation

$$(4.1) \qquad P(B \cap t^{-1}(A)) = \int_A P(B \mid t)\, dP^T(t) \quad \text{for all} \quad A \in \mathfrak{A}.$$

Suppose now that we are given a class $\mathfrak{F}$ of probability distributions for X, $\mathfrak{F} = \{P_\theta^X\}$, $\theta \in \Omega$. Denote by $P_\theta(B \mid t)$ the conditional probability of B given

$T = t$ computed for the distribution P_θ^X. The statistic T is said to be a *sufficient statistic* for $\mathfrak{F}$ (or for θ) if for every $B \in \mathfrak{B}$ there exists a determination of $P_\theta(B \mid t)$ that is independent of θ.

According to the above definition of statistic, $t(x)$ is an element of a measurable partition. However, one may consider instead any function t^* for which $t^*(x) = t^*(x')$ if and only if $t(x) = t(x')$, that is, any function that leads to this partition; the values that the function takes on are really immaterial. It will be convenient here to use this wider definition of statistic. For a rigorous treatment of some of the problems that will be referred to one needs to define an equivalence of statistics and to include in this definition the appropriate nullset considerations. A detailed account of these matters is given in [2] and [10].

From our present point of view tests are compared solely in terms of their power functions. On this basis two tests ϕ_1 and ϕ_2 may be considered equivalent if they have identical power, that is, if

$$E_\theta\phi_1(X) = E_\theta\phi_2(X) \text{ for all } \theta \in \Omega. \tag{4.2}$$

We can then state

THEOREM 4.1. *If T is a sufficient statistic for θ and $\phi(X)$ any test of a hypothesis concerning θ then there exists an equivalent test that is a function of T only.*

The proof of this theorem is immediate since

$$\psi(T) = E[\phi(X) \mid T] \tag{4.3}$$

is such a test.

It follows from Theorem 4.1 that we lose nothing by restricting consideration to tests based on a sufficient statistic.[3] The problem of determining whether or not some statistic is sufficient for a given family of distributions is simplified through the use of a criterion for sufficiency that can be checked on sight. This criterion is due to Neyman [13] who proved it in a somewhat special setting, and was recently proved in a very general form by Halmos and Savage [2]. It states that if $\mathfrak{F} = \{p_\theta\}$, $\theta \in \Omega$ is a family of generalized probability densities for X, then under certain mild restrictions a necessary and sufficient condition for $T = t(X)$ to be a sufficient statistic for $\mathfrak{F}$ is that $p_\theta(x)$ factors into one factor depending on θ but on x only through $t(x)$ and a second factor depending only on x.

The question arises as to which of various sufficient statistics to use. Since the purpose of introducing sufficient statistics is to reduce the complexity of a given statistical problem, one is led to seek a sufficient statistic that reduces the problem as far as possible and hence to the notion of a *minimal sufficient statistic*, a sufficient statistic T being *minimal* if it is a function of every other sufficient statistic (see [10]). It can be shown under fairly general conditions that a minimal sufficient statistic exists, and one can give an explicit construction for it.

---

[3] A justification for the use of sufficient statistics in the general statistical decision problem was given in [2].

As one would expect it turns out that the sufficient statistics commonly associated with various families of distributions are actually minimal. Thus for example, if $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and variance σ^2, the statistic $(\bar{X}, \Sigma(X_i - \bar{X})^2)$ is a minimal sufficient statistic for $\theta = (\xi, \sigma^2)$. If $X_1, \cdots, X_n$ are independently uniformly distributed over $(0, \theta)$, $\max(X_1, \cdots, X_n)$ is the minimal sufficient statistic for θ. If $\mathfrak{F}$ is the family of distributions according to which $X_1, \cdots, X_n$ are identically independently distributed according to an arbitrary univariate distribution (or according to an arbitrary probability density with respect to a fixed univariate measure), then the minimal sufficient statistic is obtained by defining for each point $x = (x_1, \cdots, x_n)$ the set $t(x)$ as the set of points obtainable from x by permutation of coordinates. Alternatively one can define it by $t(x_1, \cdots, x_n) = (\Sigma x_i, \Sigma x_i^2, \cdots, \Sigma x_i^n)$.

5. The principle of invariance. The notion of invariance was introduced into the statistical literature in the writings of R. A. Fisher, Hotelling, Pitman [20] and others, in connection with various special problems. A general formulation was given by Hunt and Stein who, in an unpublished paper [5], utilized this notion to find most stringent tests, and who obtained the examples of uniformly most powerful invariant tests that will be given below. The point of view in the present section is different from theirs however, since here invariance will only be considered as an intuitively appealing restriction that one may wish to impose on statistical tests.

We shall begin by considering an example. Suppose it were known that the height of people is distributed about a known mean, which for convenience we shall take to be zero, either according to a normal or to a Cauchy distribution, with unknown scale factor so that either

$$(5.1) \qquad f_\theta(x) = \frac{1}{\sqrt{2\pi}\,\theta} \exp\left(-\frac{x^2}{2\theta^2}\right) \qquad 0 < \theta < \infty$$

or

$$(5.2) \qquad f_\theta(x) = \frac{\theta}{\pi} \frac{1}{\theta^2 + x^2}, \qquad 0 < \theta < \infty.$$

Suppose we wish to test from a sample $X_1, \cdots, X_n$ the hypothesis H that the true probability density belongs to the first of these classes against the alternative that it belongs to the second. Then it seems desirable that the decision of whether or not to accept H should be independent of the scale adopted for measuring the heights. For otherwise one worker expressing his data in feet might reject H while another worker using the same data but expressing them in inches would reach the contrary decision (In this connection see for example [34], p. 104). A "nice" test function ϕ therefore would be independent of the choice of scale, i.e., it would satisfy the condition

(5.3) $\phi(cx_1, \cdots, cx_n) = \phi(x_1, \cdots, x_n)$ for all $c > 0$ and for all $(x_1, \cdots, x_n)$

except possibly on a set N, independent of c and of measure zero.

On analyzing this problem one is led to the following observation. Multiplying each of the random variables $X_1, \cdots, X_n$ by the same constant leaves both ω and $\Omega - \omega$ invariant, i.e., if the X's are normally distributed with zero mean and arbitrary scale so are $cX_1, \cdots, cX_n$, and analogously for the Cauchy distributions. It is this fact that makes it so desirable to have ϕ invariant under multiplication of the x's by a common constant.

More generally consider measurable 1:1 transformations g of $\mathfrak{X}$ into itself, and let $Y = gX$. Suppose that when X is distributed according to $\theta \in \omega$, Y is distributed according to $\theta' \in \omega$—we shall then write $\theta' = \bar{g}\theta$—and that as θ ranges over ω so does θ'. Suppose that the analogous condition is satisfied for $\Omega - \omega$, so that the problem of testing ω against $\Omega - \omega$ is left invariant under g. Now whether one expresses the observations in terms of X or in terms of Y is essentially a matter of choice of coordinates. The principle of invariance asks that if such a change of coordinates leaves the problem invariant, then it should also leave the test invariant, i.e., if G is a group of measurable 1:1 transformations of $\mathfrak{X}$ such that

$$\bar{g}\omega = \omega \text{ and } \bar{g}(\Omega - \omega) = \Omega - \omega \text{ for all } g \in G, \tag{5.4}$$

then ϕ should satisfy the condition

$$\phi(gx) = \phi(x) \text{ for all } g \in G, \tag{5.5}$$

and for all x except on a set N independent of g and such that $\mu(N) = 0$. If this condition were not satisfied, two workers, using the same data but expressing them in different coordinate systems might arrive at contrary conclusions.

As an example consider the general linear univariate hypothesis. In canonical form $X_1, \cdots, X_r; X_{r+1}, \cdots, X_s; X_{s+1}, \cdots, X_n$ are independently normally distributed with common variance. The means of the first s variables are unknown, the means of the last n-s variables are known to be zero. The hypothesis states that the first r means are zero. Adding arbitrary constants to each of the variables of the middle group leaves ω and $\Omega - \omega$ invariant. So does any orthogonal transformation of the first r variables, and any orthogonal transformation of the last $n - s$ variables. Finally, the problem is also left invariant when all of the variables are multiplied by the same constant. It is easy to see that a function ϕ is invariant under these transformations if and only if it is a function of

$$\sum_{i=1}^{r} x_i^2 \Big/ \sum_{i=s+1}^{n} x_i^2.$$

But, as is well known and easy to show, among all tests based on this statistic there is a uniformly most powerful one, namely the test that rejects H when

$$\sum_{i=1}^{r} x_i^2 \Big/ \sum_{i=s+1}^{n} x_i^2$$

is too large. Therefore, among all tests satisfying the condition of invariance the standard test is uniformly most powerful.

To formulate a corresponding reduction procedure in general, we define a function h on $\mathfrak{X}$ to be maximal invariant (under G) if it is invariant and if $h(x') = h(x)$ implies the existence of $g \in G$ such that $x' = gx$. Then a function φ on $\mathfrak{X}$ is invariant under G if and only if it depends on x only through $h(x)$, that is, if there exists a function ψ such that $\varphi(x) = \psi[h(x)]$. Hence a necessary and sufficient condition for a test to be invariant under G is that it be based on the statistic $Y = h(X)$. The principle of invariance therefore reduces the problem from X to $Y = h(X)$. To determine the resulting statistical reduction, that is, the simplification of the parameter space, one may consider the group $\bar{G}$ of transformations over Ω induced by G. If $v(\theta)$ is a maximal invariant function under $\bar{G}$, it is easily shown that the distribution of Y depends only on $v(\theta)$. Hence under the principle of invariance any two θ-values with common $v(\theta)$ (that is, such that each can be obtained from the other by a transformation of $\bar{G}$) are identified. If in particular $v(\theta)$ is constant over ω, the hypothesis H, when expressed for Y, becomes simple, and there may even exist a uniformly most powerful invariant test.

Besides for the example already mentioned this is the case for Hotelling's T^2-problem and for the hypothesis specifying the value of a multiple correlation coefficient. Another example is obtained when $X_1, \cdots, X_n$ are independently identically distributed, each with probability density $p_\theta(x)$ where under H_i $p_\theta(x) = f_i(x - \theta)$, $(i = 0, 1)$, and where it is desired to test H_0 against H_1. One may also in this example replace the location parameter by a scale parameter or have both parameters present.

It may be worth noting that the likelihood ratio test is invariant under any transformation leaving the statistical problem invariant. In the problems concerning normal distributions mentioned above, when there exists a uniformly most powerful invariant test, it coincides with the likelihood ratio test. That this is not so in general can be seen from Stein's example given in section 1. There the problem remains invariant under multiplication of X by -1, and there exists a uniformly most powerful invariant test. However, the likelihood ratio test is instead uniformly least powerful.

For certain applications it is more useful to consider a somewhat weaker definition of invariance. We shall say that a function φ is *almost invariant* under a group G of transformations if for each $g \in G$, $\varphi(gx) = \varphi(x)$ for all x except on a set N_g such that $\mu(N_g) = 0$. This definition differs from the previous one in that the null set N_g is now permitted to depend on g. It was shown by Hunt and Stein that under certain conditions on G, which are satisfied for the problems mentioned above, any almost invariant test is invariant.

We have indicated how for certain hypotheses one can find a group of transformations leaving the problem invariant, such that among all tests invariant under this group there exists a uniformly most powerful one. The question may be raised whether this approach is consistent, or whether there may exist some other group of transformations also leaving the problem invariant but leading to a different test. Also in problems where among all invariant tests there does

not exist a uniformly most powerful one, the question arises whether one is using the totality of transformations leaving the problem invariant, or whether perhaps one can reduce the problem further. It therefore seems of interest to determine the totality of transformations leaving a given problem invariant. This was carried out for a few simple problems in [8].

We finally mention a connection between the notions of invariance and sufficiency. Consider any problem in which the variables $X_1, \cdots, X_n$ are independently identically distributed under all distributions of Ω. Such a problem clearly is left invariant under any permutation of the variables. Actually, these transformations leave not only ω and $\Omega - \omega$ invariant but each point of Ω individually. No essential reduction of the problem is obtained since the maximal invariant statistic is a sufficient statistic. It is easily seen that this will always be the case when the transformations leave Ω pointwise invariant, but that in this way one does not obtain all sufficient statistics. These can be obtained, however, by considering more general transformations, where each point x of $\mathfrak{X}$ is transformed into the points of $\mathfrak{X}$ according to a probability distribution P_x.

6. The principle of unbiasedness. As a second principle of reduction we shall consider the principle of unbiasedness proposed by Neyman and Pearson. A test is said to be unbiased [19] if

$$P_\theta \text{ (rejecting } H) \geq \alpha \text{ for all } \theta \in \Omega - \omega.$$

This seems a desirable property for a test to have since it assures that there do not exist θ_0 in ω and θ_1 in $\Omega - \omega$, for which

$$P_{\theta_0} \text{ (rejecting } H) > P_{\theta_1} \text{ (rejecting } H).$$

We shall therefore be concerned in this section with the totality of tests ϕ for which

$$\begin{aligned} E_\theta \phi(X) &\leq \alpha \quad \text{for all} \quad \theta \in \omega \\ E_\theta \phi(X) &\geq \alpha \quad \text{for all} \quad \theta \in \Omega - \omega. \end{aligned} \tag{6.1}$$

For a number of important special cases there exists, among all tests satisfying (6.1), one that is uniformly most powerful in $\Omega - \omega$ and uniformly least powerful in ω. (The latter property is of course very desirable since when H is true one wants to reject it as rarely as possible.) This follows immediately from well known results concerning best similar tests since for the problems in question Ω is a subset of a Euclidean space and for any test ϕ, $E_\theta\phi(X)$ is a continuous function of θ. If then Λ is the set of points that are boundary points both of ω and of $\Omega - \omega$, it follows from (6.1) that

$$E_\theta\phi(X) = \alpha \text{ for all } \theta \in \Lambda, \tag{6.2}$$

i.e., that ϕ is similar for θ in Λ. But if among all tests satisfying (6.2) there exists one that is uniformly most powerful in $\Omega - \omega$ and uniformly least power-

ful in ω, it automatically satisfies (6.1) as is seen by comparison with the test $\phi(X) \equiv \alpha$.

As an example suppose that $X_1, \cdots, X_n$ are independently normally distributed with common mean ξ and common variance σ^2. If the hypothesis is $H_1 : \sigma \leq 1$ and the alternatives are $\sigma > 1$, the set Λ becomes the line $\sigma = 1$. As was shown by Neyman and Pearson [18], among all tests satisfying (6.2) with this Λ, the test that rejects H_1 when $\Sigma(x_i - \bar{x})^2 \leq k$ (where k is an appropriately chosen constant) is uniformly most powerful for θ in $\Omega - \omega$, and uniformly least powerful for θ in ω.

If instead we consider testing the hypothesis $H_2 : \sigma = 1$ against the alternatives $\sigma \neq 1$, we find that $\Lambda = \omega$, and our problem reduces to that of finding the best test among all those that are similar in ω and unbiased. As is well known, it turns out that rejecting when $\Sigma(x_i - \bar{x})^2 \leq k_1$ and when $\Sigma(x_i - \bar{x})^2 \geq k_2$ (where $k_1 < k_2$ are two appropriately chosen constants) is uniformly most powerful among all similar unbiased tests.

A third hypothesis concerning σ that might be of interest is $H_3 : \sigma_1 \leq \sigma \leq \sigma_2$. Here Λ consists of the two lines $\sigma = \sigma_1$ and $\sigma = \sigma_2$ and it is easy to show that the test that is uniformly most powerful in $\Omega - \omega$ and uniformly least powerful in ω rejects H_3 if and only if $\Sigma(x_i - \bar{x})^2 \leq c_1$ or $\Sigma(x_i - \bar{x})^2 \geq c_2$ where again $c_1 < c_2$ are two appropriately selected constants.

The question arises as to the connection of the principles of invariance and unbiasedness. Clearly if there exists a unique test ϕ that is uniformly most powerful unbiased, this test is invariant under any group G leaving the problem invariant. If then in addition there exists a uniformly most powerful invariant (under G) test, this must coincide with ϕ. Thus, if both principles lead to a unique optimum solution, these solutions coincide.

We have seen that frequently optimum unbiased tests can be obtained through a study of tests that are similar over certain sets in the parameter space. The totality of similar tests was obtained for a number of important problems by Neyman and Pearson. In his 1937 paper on confidence intervals [15] Neyman gave a general method for constructing similar regions with the help of sufficient statistics. Let T be a sufficient statistic for $\theta \in \Lambda$. The condition for ϕ to be similar with respect to Λ and of size α, is that

$$E_\theta\phi(X) = E_\theta E[\phi(X) \mid T] = \alpha \text{ for all } \theta \in \Lambda, \tag{6.3}$$

i.e., that

$$E_\theta\{E[\phi(X) \mid T] - \alpha\} = 0 \text{ for all } \theta \in \Lambda. \tag{6.4}$$

Clearly any test ϕ for which

$$E[\phi(X) \mid t] = \alpha \text{ for almost all } t \tag{6.5}$$

is similar. This is the construction given by Neyman, and we shall say that a test ϕ satisfying (6.5) has the Neyman structure with respect to T. The question whether this exhausts the totality of similar tests is easily reduced to an

analytic problem the solution of which is known in many special cases. This method was first employed by P. L. Hsu [3] for some problems concerning normal distributions, and was extended to other cases in [7]. The present general formulation was given by H. Scheffé and the author in [9] and [10]. We shall say that a family of distributions $\{P_\theta^T\}$, $\theta \in \Lambda$, is boundedly complete if

(i) $f(t)$ is bounded,

(ii) $E_\theta f(T) = 0$ for all $\theta \in \Lambda$

imply $f(t) = 0$ except on a set N with $P_\theta(N) = 0$ for all $\theta \in \Lambda$. Then we can state

THEOREM 6.1. *A necessary and sufficient condition for the totality of tests similar for Λ to have Neyman structure with respect to a sufficient statistic T is that $\{P_\theta^T\}$, $\theta \in \Lambda$, be boundedly complete.*

7. Tests whose power increases with the distance from the hypothesis. Frequently, even among the unbiased tests, there does not exist a uniformly most powerful one. The general univariate linear hypothesis with more than one constraint is an example of this situation. The following extension of the idea of unbiasedness may then be used to reduce the class of tests still further. Unbiasedness distinguishes between values of θ as they belong to ω or $\Omega - \omega$. However, one may further classify the points of $\Omega - \omega$ according to their "distance" from ω, and then ask of a test φ that the further be θ from ω the larger be the power $\beta_\varphi(\theta)$.

One possible such ordering of the alternatives is that induced by the envelope power function. Here the envelope power at θ (Wald [24]) is defined by

$$\beta_\alpha^*(\theta) = \sup_{\varphi \in \mathfrak{F}(\alpha)} \beta_\varphi(\theta) \tag{7.1}$$

where $\mathfrak{F}(\alpha)$ is the class of all tests φ with $E_\theta\varphi(X) \leq \alpha$ for all $\theta \in \omega$. Of two points θ_1, θ_2 one may then say that θ_1 is closer to ω than θ_2, equally close or less close, as $\beta_\alpha^*(\theta_1)$ is less than, equal to or greater than $\beta_\alpha^*(\theta_2)$. The distance of θ from ω is thus measured by the ease with which one can detect that the hypothesis is false when θ is the true parameter value.

When θ lies in a Euclidean space and $\beta_\varphi(\theta)$ is a continuous function of θ for all θ, as is the case in most applications, the condition that the power increase with β_α^* will usually imply that $\beta_\varphi(\theta_1) = \beta_\varphi(\theta_2)$ whenever $\beta_\alpha^*(\theta_1) = \beta_\alpha^*(\theta_2)$. In the case of the general linear hypothesis considered in section 5, for example, one would obtain the condition that the power be a function only of $\sum_{i=1}^{r} \xi_i^2/\sigma^2$ where $\xi_i = E(X_i)$. As was shown by P. L. Hsu [3], the standard (likelihood ratio) test is uniformly most powerful among all tests satisfying this condition. Analogous remarks apply to Hotelling's T^2- problem, and to the hypothesis specifying the value of the multiple correlation coefficient. The corresponding optimum properties in these cases were proved by Simaika [21].

It is interesting to compare the above condition with that of invariance.

This comparison yields nothing of interest if the totality of tests is considered. We may, however, restrict our attention to tests depending only on a sufficient statistic T. We already know that $\varphi(X)$ and $E[\varphi(X) \mid T]$ have identical power. In order to validate the comparison we wish to make, we state the following

LEMMA. *Let T be a sufficient statistic for $\theta \in \Omega$, and let G be a group of 1:1 transformations g on X leaving Ω invariant. Then if $\varphi(x)$ is invariant under G, $E[\varphi(X) \mid t]$ is almost invariant under G.*

We can now state the desired comparison in the following

THEOREM 7.1. *Let G be a group of 1:1 transformations on X, let $\bar{G}$ be the induced group of transformations on Ω, let $v(\theta)$ be maximal invariant under $\bar{G}$, and suppose that $\bar{G}$ leaves ω and $\Omega - \omega$ invariant. Suppose further that T is a sufficient statistic for Ω, and that $\{P_\theta^T\}$, $\theta \in \Omega$, is boundedly complete. Then a necessary and sufficient condition that the power of a test $\psi(T)$ be a function only of $v(\theta)$, is that $\psi(t)$ be almost invariant under G.*

This theorem is an immediate extension of some results of Wolfowitz [35].

Theorem 7.1 together with the results of section 5 proves that the standard tests of the general linear hypothesis, Hotelling's T^2-problem and the hypothesis concerning the multiple correlation coefficient possess the optimum property that was obtained for these problems by Hsu and Simaika, respectively. The method of proof indicated here is due to Wolfowitz [35].

8. Most stringent tests. We shall now turn to the third aspect of the theory: Optimum properties defined with reference to the whole class of alternatives, and attainable with no restrictions imposed on the class of tests. In the present section we shall consider the property of stringency. Wald [25] defines a test φ to be most stringent if it minimizes

$$\sup_{\theta \in \Omega - \omega} [\beta_\alpha^*(\theta) - \beta_\varphi(\theta)], \tag{8.1}$$

where β_α^* again denotes the envelope power, and β_φ the power of φ. The rationale of this definition is clear. The difference $\beta_\alpha^*(\theta) - \beta_\varphi(\theta)$ measures the amount by which the test falls short at the alternative θ of the power that could be attained against this particular alternative. A test φ is therefore most stringent if it minimizes its maximum shortcoming.

A theory of most stringent tests was developed by Hunt and Stein [5], who based it on the notion of invariance. Consider, as in section 5, a group G of measurable 1:1 transformations on $\mathfrak{X}$ leaving the problem invariant. Hunt and Stein obtained their results in connection with the following groups of transformations.

(i) $gx = x + c$, $-\infty < c < \infty$, x a real variable;
(ii) $gx = ax$, $0 < a$, x a real variable;
(iii) $gx = ax + c$, $0 < a$, $-\infty < c < \infty$, x a real variable;
(iv) the group of orthogonal transformations on a Euclidean space;
(v) any finite group.

THEOREM 8.1. (Hunt and Stein). *If G is the direct product of a finite number of groups of types* (i)–(v), *and if G leaves the problem invariant, that is, if G satisfies* (5.4), *then there exists a most stringent test invariant under G.*

Actually, it is not necessary here to require that G be a direct product. The result holds also if the factoring of G is according to normal subgroups, where the normal subgroup at each stage and the final factor group are of the types mentioned. In the light of this one may omit type (iii) from the list since it has a normal subgroup of type (i) with factor group of type (ii).

The proof of Theorem 8.1 is based on the following lemma, which has applications to many related problems.

LEMMA (Hunt and Stein). *If G is a direct product of a finite number of groups of types* (i)–(v) *then given any function f over $\mathfrak{X}$ $(0 \leq f(x) \leq 1)$ there exists a function F $(0 \leq F(x) \leq 1)$ such that F is invariant under G, and*

$$(8.2) \qquad \inf_{g \in G} \int f(gx)\varphi(x)\, d\mu(x) \leq \int F(x)\varphi(x)\, d\mu(x) \leq \sup_{g \in G} \int f(gx)\varphi(x)\, d\mu(x)$$

for all φ that are integrable μ.

It follows from Theorem 8.1 that if there exists a uniformly most powerful invariant test, this test is most stringent. In this way Hunt and Stein show, for example, (see in this connection section 5), that the likelihood ratio test of the general univariate linear hypothesis is most stringent. A question that is left open is the uniqueness of such a most stringent test.

In general, the possibility therefore remains that there might exist another most stringent test uniformly more powerful than the invariant one. In certain particular cases this possibility can be ruled out by the following consideration. Suppose that Ω is a subset of a Euclidean space and that every point of ω is a limit point of $\Omega - \omega$. Suppose further that for any test ϕ, $E_\theta\phi(X)$ is continuous in θ. Then clearly, if ϕ_1 is similar of size α for testing ω and ϕ_2 is of size $\leqq \alpha$ but not similar, ϕ_2 can not be uniformly as powerful as ϕ_1. Hence any test that is admissible among all similar tests of size α is also admissible among the totality of tests of size $\leqq \alpha$. Now admissibility among all similar tests is sometimes not too difficult to prove. For the likelihood ratio test of the general linear univariate hypothesis, for example, it is an immediate consequence of the properties of this test proved by Wald [23] and Hsu [4].

The following alternative method for obtaining most stringent tests is also mentioned by Hunt and Stein.

THEOREM 8.2. (Hunt and Stein). *Let $\Omega - \omega$ be partitioned into disjoint subsets Ω_δ such that $\beta^*_\alpha(\theta)$ is constant on each Ω_δ, and let φ_δ be the test that maximizes $\inf_{\theta \in \Omega_\delta} \beta_{\varphi\delta}(\theta)$. Then if $\varphi_\delta = \varphi$ is independent of δ, φ is most stringent.*

This result may be supplemented by the following method for finding tests that maximize $\inf_{\theta \in \omega_1} \beta_\varphi(\theta)$ over a given set of alternatives ω_1 (not necessarily satisfying the conditions imposed above on the Ω_δ's).

THEOREM 8.3. *Suppose additive classes of sets have been defined over ω and ω_1, and consider probability measures λ and λ_1 over ω and ω_1. Let the functions $f_\theta(x)$ be generalized probability densities with respect to μ, so that $h(x) = \int_\omega f_\theta(x)\, d\lambda(\theta)$ and $h_1(x) = \int_{\omega_1} f_\theta(x)\, d\lambda_1(\theta)$ are again probability densities with respect to μ. Let φ be the most powerful test of size α for testing the simple hypothesis $H{:}\,h$ against the simple alternative h_1, and suppose that the power of φ against h_1 is β. Then if*

$$
\begin{aligned}
&E_\theta\varphi(x) \leq \alpha \quad \textit{for all} \quad \theta \in \omega, \\
&E_\theta\varphi(x) \geq \beta \quad \textit{for all} \quad \theta \in \omega_1,
\end{aligned} \tag{8.3}
$$

φ maximizes $\inf_{\theta \in \omega_1} \beta_\varphi(\theta)$ at level of significance α.

This method, when applicable, has the advantage of giving the totality of most stringent tests (see in this connection Theorem 3.1) and hence of settling the question of admissibility. However, in many applications probability measures λ, λ_1 with the desired properties do not exist but instead only sequences $\lambda^{(n)}$, $\lambda_1^{(n)}$, which satisfy the conditions in the limit. In this case again only the weak conclusion is possible: The test obtained is most stringent but has not been proved admissible. (For an example in which the analogous method has been carried through in detail for an estimation problem, see [22]).

Actually, the two methods are closely related, as can be seen from the proof of the main lemma. In those cases in which there exists a group G giving the maximum possible reduction, the group $\bar{G}$ induces a partition of Ω (through the equivalence: $\theta_1 \sim \theta_2$ if there exists $\bar{g}$ such that $\theta_2 = \bar{g}\theta_1$), just into ω and the sets Ω_δ. (This is so mainly because, as was shown by Hunt and Stein, the envelope power remains invariant under any transformations that leave the problem invariant.) Then the measures λ, λ_δ over ω, Ω_δ respectively, which figure in the application of Theorems 8.2 and 8.3, become invariant measures over $\bar{G}$ through the obvious 1:1 mapping from ω and the Ω_δ's respectively to $\bar{G}$. Thus the second method will allow the strong conclusion when the group $\bar{G}$ involved in the first method possesses a finite invariant measure [types (iv) and (v)] but not if any of its factors are of type (i)–(iii).

To conclude this section we shall give an example where the method of invariance leads only to a partial reduction but where the solution may be completed by certain additional considerations. Suppose that $(X_1, \cdots, X_n)$ is a sample from a normal distribution with mean ξ and variance σ^2, both unknown, and that we wish to find the most stringent test of the hypothesis H: $\sigma = 1$ against the alternatives $\sigma \neq 1$. Theorem 8.1 reduces the problem to the statistic $Y = \Sigma(X_i - \bar{X})^2$, but among the tests of H based on this statistic there does not exist a uniformly most powerful one. It may also be shown [8] that no further reduction is possible by means of the method of invariance.

However, one may now consider the problem of finding the most stringent test based on Y. (The envelope power function $\beta^*(\xi, \sigma)$ that must be used

naturally is not the one for Y but that for the original problem.) From an argument given in [6] it follows that this test is of the form

$$\varphi_{k_1,k_2}\text{: reject when } Y < k_1 \text{ or } > k_2 ,$$

where k_1, k_2 are determined by the two conditions

(i) $P(\text{rejection} \mid \sigma = 1) = \alpha$,

(ii) $\sup_{\sigma<1} [\beta_\alpha^*(\xi, \sigma) - \beta_{\varphi_{k_1,k_2}}(\sigma)] = \sup_{\sigma>1} [\beta_\alpha^*(\xi, \sigma) - \beta_{\varphi_{k_1,k_2}}(\sigma)]$.

Here $\beta_\alpha^*(\xi, \sigma)$ is independent of ξ and can be obtained from a table of the χ^2-distribution (with n degrees of freedom for $\sigma < 1$ and n-1 degrees of freedom for $\sigma > 1$ as can be seen from (3.6)). Hence k_1 and k_2 can be computed fairly easily.

Another problem that may be treated in this way is the hypothesis of equality of variances for two normal samples. If the two samples are of equal size, there exists a uniformly most powerful invariant test for a suitable group of transformations. However, if the sample sizes are different the method of invariance reduces the problem only to $\Sigma(X_i - \bar{X})^2 / \Sigma(Y_i - \bar{Y})^2$, and the cut off points giving the most stringent test may be determined by an argument analogous to that given above.

This method may be extended to allow determination of most stringent test of hypotheses such as H: $\sigma_1 \leqq \sigma \leqq \sigma_2$. This requires a certain modification of Theorem 1 of [6], which is easily obtained. One finds agains that one may restrict consideration to a one-parameter family of tests (determined by a somewhat different condition than above), and that among these the most stringent test is obtained by the analogue of condition (ii) above.

If should be mentioned that the results of [6] apply also to the hypothesis specifying the value of the parameter in a binomial or Poisson distribution. This is easily seen since in either case the distributions of Ω are absolutely continuous with respect to a common sigma finite measure and since for the appropriate choice of this measure the generalised density is of the form assumed for the density in [6]. Hence in both the binomial and the Poisson case the most stringent test is determined by conditions analogous to (i) and (ii) above.

9. Tests that minimize the maximum loss. In the Neyman-Pearson theory one classifies the errors into two kinds: Rejecting the hypothesis when it is true, accepting it when it is false. One may however analyze the situation further and distinguish, say, between accepting when one or some other alternative is true. Thus one is led to introduce the losses that result in a given situation from the various possible errors, and to look for a test that, in an appropriate sense, minimizes the expected loss. This possibility was mentioned by Neyman and Pearson [17], and was taken as the starting point of his general theory by Wald (see for example [24]).

In order to stay within the framework of this exposition we shall here introduce losses only for the errors of accepting the hypothesis when it is false,

while still demanding that the probability of rejection when the hypothesis is true should not exceed α. Actually, there are many cases where this seems to be a reasonable formulation. For it frequently happens that the two types of error entail consequences of such completely different nature that the resulting losses cannot be measured on a common scale while usually the different errors of the same type are comparable.

We shall therefore assume that for each $\theta \in \Omega - \omega$ there is defined a $W(\theta)$, which measures the loss resulting from acceptance of H when θ is true. The risk which one runs by using a test φ, when $\theta \in \Omega - \omega$ is the true parameter value is given by the expected loss $R_\varphi(\theta) = W(\theta)\, E_\theta[1 - \varphi(X)]$. When a uniformly most powerful test exists for the hypothesis in question, this test also minimizes the expected loss uniformly for θ in $\Omega - \omega$. In the contrary case one may again restrict the class of tests in some way, so that within the restricted class there exists a uniformly most powerful test, and hence a test that uniformly minimizes the expected loss. Alternatively we may again consider some optimum property of the risk function $R_\varphi(\theta)$ as a whole. We shall here consider the minimax principle introduced by Wald, and seek a test, which, subject to $E_\theta\varphi(X) \leq \alpha$ for all $\theta \in \omega$, minimizes

$$\sup_{\theta \in \Omega - \omega} W(\theta)\, E_\theta[1 - \varphi(X)],$$

the maximum risk.

If one introduces losses also for the other type of error it is easy to see that for a suitably chosen loss function the definition of minimax expected loss coincides with that of stringency. It is therefore not surprising that the methods of the previous section can be extended to cover the problems considered in the present one. (They are actually much more general, and may be applied also, for example, to the problem of point estimation, and in fact to the general decision problem).

From the lemma of Hunt and Stein stated in the previous section we immediately obtain the following extension of Theorem 8.1.

THEOREM 9.1. *If G is a group of transformations leaving the hypothesis and the class of alternatives invariant, if G can be factored by normal subgroups into factors of types* (i)–(v), *and if the loss function $W(\theta)$ is invariant under $\bar{G}$, then there exists a test φ invariant under G and minimizing*

$$\sup_{\theta \in \Omega - \omega} W(\theta)\, E_\theta[1 - \varphi(X)]. \tag{9.1}$$

It follows that when a uniformly most powerful invariant test exists, this test has the property of minimizing the maximum expected loss with respect to any invariant loss function. Thus Student's test, for example, minimizes the maximum risk for any loss function that depends only on $|\xi|/\sigma$.

Clearly the second method mentioned in section 8 can be extended in an analogous manner if in Theorem 8.2 one replaces the sets Ω_δ by sets over which $W(\theta)$ is constant.

Again it may happen that the method of invariance does not reduce the problem sufficiently far but that the solution may be completed by other considerations. Let us once more consider the hypothesis $H: \sigma = 1$ of the previous section, and let us suppose that the loss function has the necessary invariance property, so that it is a function only of σ but not of the unknown mean. It follows from Theorem 9.1 that there exists a test minimizing the maximum risk, which is a function only of $Y = \Sigma(X_i - \bar{X})^2$. From [6] it is easily seen that a test φ_{k_1,k_2} which rejects when $Y < k_1$ or $> k_2$, has the desired property if its size is α and if in addition

$$\sup_{\sigma<1} W(\sigma)E_\sigma[1 - \varphi(Y)] = \sup_{\sigma>1} W(\sigma)E_\sigma[1 - \varphi(Y)]. \tag{9.2}$$

It follows that depending on the choice of $W(\sigma)$ the solution may be any member of the one-parameter family of tests φ_{k_1,k_2} of size α.

Under the conditions of Theorem 9.1, when a uniformly most powerful invariant test exists, this also maximizes the average power for a large class of weight functions. If there exists a common finite invariant measure over the sets Ω_δ in the sense indicated in section 8, the uniformly most powerful invariant test will maximize the average power with this measure as weight function, over Ω_δ for all δ. It follows that it maximizes the average power over $\Omega - \omega$ with respect to any weight function for which the conditional distribution over each Ω_δ is the above invariant measure. If the invariant measure over the Ω_δ's is not finite one can obtain analogous results with respect to a sequence of weight functions invariant in the limit. The results indicated here are much weaker than those obtained for the general linear univariate hypothesis by Wald [23] and Hsu [4] under the restriction to similar regions. However their results are no longer valid when this restriction is omitted.

10. Applications to sequential analysis. So far we have restricted consideration to the case that the hypothesis is to be tested on the basis of a preassigned experiment. However, frequently there is available for this purpose a large class of experiments, and the selection of an optimum experiment out of this class is part of the problem. We shall consider here only the following situation, which has recently been studied extensively (see Wald [28, 29]). There is given a sequence of random variables $X_1, X_2, \cdots$ whose joint distribution is known to belong to some family $\mathfrak{F} = \{P_\theta\}$, $\theta \in \Omega$; the hypothesis specifies some subfamily: $\theta \in \omega$. The X's are observed one by one, and the decision, whether or not to continue experimentation at any given stage, is allowed to depend on the observations taken up to that point. Thus the number n of observations that will be taken is a random variable whose distribution depends on θ. Usually, by an appropriate choice of stopping rule, there may be effected a considerable saving in the expectation of the number of observations necessary to achieve a given discrimination between hypothesis and alternatives. The problem is to determine the stopping rule and test that minimizes this expectation.

As we have seen in the previous sections the principal methods for obtaining

optimum tests consist in reducing the problem to that of testing a simple hypothesis against a simple alternative. This basic problem was solved in the non-sequential case by Neyman and Pearson (Theorem 3.1). The solution of the much more difficult corresponding sequential problem was obtained for a large class of cases by Wald and Wolfowitz [31] in the following

THEOREM 10.1. *Let $X_1, X_2, \cdots$ be identically and independently distributed. It is desired to test the hypothesis that the common probability density of the X's is $f(x)$ against the alternative that it is $g(x)$. Given two numbers $0 < \alpha < \beta < 1$, there exists a test which, subject to the condition*

$$\begin{aligned} &P\,(\text{rejection} \mid f) \leq \alpha \\ &P\,(\text{rejection} \mid g) \geq \beta, \end{aligned} \tag{10.1}$$

minimizes simultaneously $E_f(n)$ and $E_g(n)$, the expected number of observations computed for the distributions f and g. This test is given in terms of two numbers A and B by the following rule. After m observations have been taken,

$$\textit{reject if } \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} > A,$$

$$\textit{accept if } \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} < B,$$

$$\textit{take another observation if } B < \frac{g(x_1) \cdots g(x_m)}{f(x_1) \cdots f(x_m)} < A.$$

Here A and B are determined so that condition (10.1) *holds with the inequality signs replaced by equality.*

So as to be able to treat the various problems considered non-sequentially in the previous sections one would have to extend this theorem at least to the case that the variables $X_1, X_2, \cdots$ form a set of equivalent variables in the sense of de Finetti [1]. Instead, we shall here restrict ourselves to a few problems that can be solved on the basis of Theorem 10.1. All of the tests discussed below were derived from various points of view and some of their properties were discussed by Girshick in his important "Contributions to the theory of sequential analysis", *Annals of Math. Stat.*, vol. 17 (1946) pp. 123–143 and 282–298, and by Wald in his basic book on the subject [28].

It is convenient here to modify slightly the formulation of the problem of hypothesis testing. Let the parameter space Ω be divided into three sets, the set ω_0 specified by the hypothesis, the class of alternatives ω_1, and a region of indifference $\Omega - \omega_0 - \omega_1$ where we do not much care whether the hypothesis is accepted or rejected (see [28]). Let us denote the sequential random variable $(X_1, \cdots, X_n)$ by X. Then we wish to determine a sequential test φ, which, subject to

$$\begin{aligned} &E_\theta\varphi(X) \leq \alpha \text{ for } \theta \in \omega_0 \\ &E_\theta\varphi(X) \geq \beta \text{ for } \theta \in \omega_1, \end{aligned} \tag{10.2}$$

minimizes $\sup_{\theta \epsilon \omega_0 + \omega_1} E_\theta(n)$. (Actually, this is a rather artificial formulation. The natural requirement is the minimization of $\sup_{\theta \epsilon \Omega} E_\theta(n)$ but this is a much more difficult problem.) The reduction to the problem of testing a simple hypothesis against a simple alternative is achieved by the following obvious extension of Theorem 8.3.

THEOREM 10.2. *Let λ_0, λ_1 be distributions over ω_0, ω_1 respectively, and let φ be a test, which subject to*

$$\int_{\omega_0} E_\theta \varphi(X) d\lambda_0(\theta) \leq \alpha$$
$$(10.3) \qquad \int_{\omega_1} E_\theta \varphi(X) d\lambda_1(\theta) \geq \beta,$$

minimizes $\sup_{i \epsilon \{0,1\}} \int E_\theta(n)\, d\lambda_i(\theta)$. *Then if φ satisfies* (10.2) *and*

$$(10.4) \qquad E_\theta(n) \leq \sup_{i \epsilon \{0,1\}} \int E_\theta(n)\, d\lambda_i(\theta) \text{ for all } \theta \epsilon \omega_0 + \omega_1,$$

φ minimizes $\sup_{\omega_1 + \omega_2} E_\theta(n)$ *subject to* (10.2).

As in section 3 we can make certain trivial applications to problems concerning a single real parameter such as testing the hypothesis $H: p \leq p_0$ against the alternatives $p \geq p_1$ $(p_0 < p_1)$, where p is the probability of success in a binomial sequence of trials. In this example condition (10.2) of Theorem 10.2 obviously is satisfied when λ_0 and λ_1 assign probability 1 to p_0 and p_1 respectively. Hence the probability ratio test for testing $p = p_0$ against $p = p_1$ has the desired properties, whenever (10.4) holds, that is, whenever $E_p(n)$ attains its maximum between p_0 and p_1.

The following is another example that may be solved in this manner. Let $X_1, X_2, \cdots; Y_1, Y_2, \cdots$ be independently normally distributed, all with unit variance and means $E(X_i) = \xi$, $E(Y_i) = \eta$. In order to test the hypothesis $H: \xi \geq \eta$ against the alternatives $\eta - \xi \geq \delta$ where $\delta > 0$ is given, a pair (X_1, Y_1) is observed. If after this observation experimentation continues another pair (X_2, Y_2) is observed, etc. In this case we may take for λ_0, λ_1 the distributions that assign probability 1 to the parameter points $(\xi, \eta) = (0, 0)$ and $\left(-\frac{\delta}{2}, \frac{\delta}{2}\right)$ respectively. Then the probability ratio after m observations is given by

$$(10.5) \qquad \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\left(x_i + \frac{\delta}{2}\right)^2 - \frac{1}{2}\sum_{i=1}^{m}\left(y_i - \frac{\delta}{2}\right)^2\right]}{e^{-\frac{1}{2}\Sigma x_i^2 - \frac{1}{2}\Sigma y_i^2}} = e^{-(m\delta^2/4) + \delta[\Sigma y_i - \Sigma x_i]}.$$

Since the distribution of $Y - X$ depends only on $\eta - \xi$, it is easily seen that condition (10.2) is satisfied.

Some further results can be obtained through extension to the sequential case of Theorems 8.1 and 9.1.

THEOREM 10.3. *Suppose that G is of the type described in Theorem 9.1, let $Y = f(X_1, X_2, \cdots)$ be maximal invariant under G, let $v(\theta)$ be maximal invariant under $\bar{G}$, and let the set of values of $v(\theta)$ corresponding to ω_0 and ω_1 be $\bar{\omega}_0$ and $\bar{\omega}_1$, respectively. If among all tests of $\bar{\omega}_0$ against $\bar{\omega}_1$ based on Y, the test φ minimizes $\sup_{v(\theta)\epsilon\bar{\omega}_0+\bar{\omega}_1}$ - $E_\theta(n)$ subject to*

$$\text{(10.6)} \qquad \begin{aligned} E_\theta\varphi(Y) &\leq \alpha \text{ if } v(\theta)\epsilon\bar{\omega}_0 \\ E_\theta\varphi(Y) &\geq \beta \text{ if } v(\theta)\epsilon\bar{\omega}_1, \end{aligned}$$

then φ also minimizes $\sup_{\omega_0+\omega_1} E_\theta(n)$ among all tests based on the X's and which satisfy (10.2).

As an example consider the problem of testing the hypothesis $\sigma \leq \sigma_0$ against the alternatives $\sigma \geq \sigma_1$ $(\sigma_0 < \sigma_1)$ when the X's are identically, independently normally distributed with unknown mean and variance. Since the problem remains invariant under a common translation of the X's we can take for Y of the theorem $Y = (X_2 - X_1, X_3 - X_1, \cdots)$. Equivalently we may take as our new sequence of variables $(Y_1, Y_2, \cdots)$ where

$$\text{(10.7)} \qquad Y_k = \frac{kX_{k+1} - (X_1 + \cdots + X_k)}{\sqrt{k(k+1)}}.$$

Then $Y_1, Y_2, \cdots$ are independently normally distributed with zero mean and the same variance as the X's. Hence the problem reduces to a type which we have already considered. The optimum test is based on

$$\sum_{i=1}^{m} Y_i^2 = \sum_{i=1}^{m+1}\left(X_i - \frac{X_1 + \cdots + X_{m+1}}{m+1}\right)^2.$$

It may be worth pointing out that Theorems 3.2, 8.3, 10.2 all are special cases of simple results in the general theory of statistical decision functions, of which the following is the prototype. (For a detailed treatment of this theory see, for example, [30]). Let $\{P_\theta\}$, $\theta\epsilon\Omega$, be the family of possible distributions of a random variable X, and let $\{\delta\}$ be a family of decision functions. The loss resulting from the use of $\delta(x)$ when P_θ is the true distribution is $W[\theta, \delta(x)]$ and the risk function associated with δ is $R_\delta(\theta) = E_\theta W[\theta, \delta(X)]$. Let λ be a probability measure over Ω, and let δ_λ be a decision function that minimizes $\int R_\delta(\theta)\, d\lambda(\theta)$. Then if λ is such that

$$\text{(10.8)} \qquad R_{\delta_\lambda}(\theta) \leqq \int R_{\delta_\lambda}(\zeta)\, d\lambda(\zeta) \text{ for all } \theta \epsilon \Omega,$$

δ_λ minimizes $\sup_\theta R_\delta(\theta)$.

PROOF. Let δ^* be any other decision function. Then

$$\sup_\theta R_{\delta_\lambda}(\theta) \leq \int R_{\delta_\lambda}(\theta)\, d\lambda(\theta) \leqq \int R_{\delta^*}(\theta)\, d\lambda(\theta) \leqq \sup_\theta R_{\delta^*}(\theta).$$

In an analogous manner one can give an extension of Theorems 8.1, 9.1, 10.3.

11. Two sided tests considered as 3-decision problems. In a number of important special problems the hypothesis specifies the value of a real valued parameter or states that this parameter lies in a certain interval, and it is desired to test this hypothesis against the obvious two-sided class of alternatives. It seems that in nearly any problem of this kind that would arise in practice one would want to decide when rejecting the hypothesis, whether the true parameter value lies below or above the hypothetical ones. If for example one rejects the hypothesis that the means of two normal populations are equal, one usually wants to decide which of the two is larger. It would therefore seem most natural to formulate such problems as 3-decision problems.

Problems of this kind, as all problems of hypothesis testing, naturally are special cases of the general decision problem formulated by Wald. We shall here consider the case that upper bounds are given for the probabilities of certain types of errors and thereby obtain a formulation, which is closely analogous to the classical formulation of hypothesis testing discussed in this paper, and which will allow immediate application of a large portion of the theory discussed here.

Consider the case that Ω is partitioned into 3 parts, ω, ω_1, ω_2 where in a certain sense ω lies between ω_1 and ω_2. We wish to test the hypothesis $H: \theta \epsilon \omega$. When we reject the hypothesis, we shall reach either decision D_1 that $\theta \epsilon \omega_1$ or decision D_2 that $\theta \epsilon \omega_2$. Correspondingly we prescribe two positive numbers α_1, α_2 and impose the restriction that

$$(11.1) \qquad \begin{aligned} P_\theta(D_1) &\leq \alpha_1 \text{ if } \theta \epsilon \omega + \omega_2 \\ P_\theta(D_2) &\leq \alpha_2 \text{ if } \theta \epsilon \omega + \omega_1 . \end{aligned}$$

Subject to this condition it is desired to maximize

$$(11.2) \qquad \begin{aligned} &P_\theta(D_1) \text{ for } \theta \epsilon \omega_1 \\ &P_\theta(D_2) \text{ for } \theta \epsilon \omega_2 . \end{aligned}$$

A test will now consist of two non-negative functions ϕ_1 and ϕ_2 satisfying

$$(11.3) \qquad \phi_1(x) + \phi_2(x) \leq 1,$$

with the convention that when $X = x$ the decision D_i will be taken with probability $\phi_i(x)$ $(i = 1, 2)$.

There is no difficulty concerning the extension of the notions of invariance or sufficient statistic, in fact these notions obviously apply to the general decision problem. The notion of unbiasedness is extended in the obvious way by the condition

$$(11.4) \qquad \begin{aligned} P_\theta(D_1) &\geq \alpha_1 \text{ for } \theta \epsilon \omega_1 \\ P_\theta(D_2) &\geq \alpha_2 \text{ for } \theta \epsilon \omega_2 . \end{aligned}$$

One then obtains the following

THEOREM 11.1. *Suppose that for testing the hypothesis H_1: $\theta \epsilon \omega + \omega_2$ against the alternatives $\theta \epsilon \omega_1$ at level of significance α_1, the test ϕ_1 among all unbiased tests*

is uniformly most powerful in $\omega + \omega_2$ *and uniformly least powerful in* ω_1, *and that* ϕ_2 *has the analogous property for testing* $H_2 : \theta \epsilon \omega + \omega_1$ *against* $\theta \epsilon \omega_2$ *at significance level* α_2. *If* $\phi_1(x) + \phi_2(x) \leqq 1$ *for all* x, *then among all procedures satisfying* (11.1) *and* (11.4), *the procedure* (ϕ_1, ϕ_2) *uniformly maximizes the probability of a correct decision.* (If the tests ϕ_1, ϕ_2 take on only the values 0 and 1, the condition $\phi_1(x) + \phi_2(x) \leqq 1$ states that the rejection region of each of the two hypotheses is contained in the acceptance region of the other.)

As an example consider the case that $X_1, \cdots, X_n$ are independently, normally distributed with common mean ξ and variance σ^2. Suppose we wish to test the hypothesis that $\sigma_1 \leqq \sigma \leqq \sigma_2$ where σ_1 may equal σ_2. Then it follows from Theorem 11.1 that among all unbiased procedures of level (α_1, α_2), there exists one that maximizes the probability of a correct decision uniformly in ξ, σ. This is the procedure under which decision D_1 or D_2 is taken as $\Sigma(x_1 - \bar{x})^2 \leqq k_1$ or $\geqq k_2$ and the hypothesis is accepted otherwise. Here the k's are determined by

$$P(\Sigma(x_i - \bar{x})^2 \leqq k_1 \mid \sigma_1) = \alpha_1 \tag{11.5}$$
$$P(\Sigma(x_i - \bar{x})^2 \geqq k_2 \mid \sigma_2) = \alpha_2 .$$

REFERENCES

[1] B. de Finetti, "La prévision: ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré*, Vol. 7 (1937), p. 1.
[2] P. R. Halmos and L. J. Savage, "Applications of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), p. 225.
[3] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), p. 62.
[4] P. L. Hsu, "On the power function of the E^2-test and the T^2-test," *Annals of Math. Stat.*, Vol. 16 (1945), p. 278.
[5] G. Hunt and C. Stein, "Most stringent tests of statistical hypotheses," unpublished.
[6] E. L. Lehmann, "On families of admissible tests," *Annals of Math. Stat.*, Vol. 18 (1947), p. 97.
[7] E. L. Lehmann, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), p. 473.
[8] E. L. Lehmann, "On the totality of transformations leaving a statistical problem invariant," unpublished.
[9] E. L. Lehmann and H. Scheffé, "On the problem of similar regions," *Proc. Nat. Acad. Sci.*, Vol. 33 (1947), p. 382.
[10] E. L. Lehmann and H. Scheffé, "Completeness, similar regions and unbiased estimation," unpublished.
[11] E. L. Lehmann and C. Stein, "Most powerful tests of composite hypotheses. I. Normal distributions," *Annals of Math. Stat.*, Vol. 19 (1948), p. 495.
[12] E. L. Lehmann and C. Stein, "On the theory of some non-parametric hypotheses," *Annals of Math. Stat.*, Vol. 20 (1949), p. 28.
[13] J. Neyman, "Su un teorema concernente le cosidette statistiche sufficienti," *Giorn. Ist. Ital. Att.*, Vol. 6 (1935), p. 320.
[14] J. Neyman, "Sur la vérification des hypothèses statistiques composées," *Bull. Soc. Math. France*, Vol. 63 (1935), p. 1.
[15] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), p. 333.
[16] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference," *Biometrika*, Vol. 20A (1928), p. 175 and p. 263.

[17] J. Neyman and E. S. Pearson, "On the testing of statistical hypotheses in relation to probability a priori," *Proc. Camb. Phil. Soc.*, Vol. 29 (1933), p. 492.
[18] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 231 (1933), p. 289.
[19] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of type A and type A_1," *Stat. Res. Mem.*, Vol. 1 (1936), p. 1.
[20] E. J. G. Pitman, "Tests of hypotheses concerning location and scale parameters," *Biometrika*, Vol. 31 (1939), p. 200.
[21] J. B. Simaika, "On an optimum property of two important statistical tests," *Biometrika*, Vol. 32 (1941), p. 70.
[22] C. Stein and A. Wald, "Sequential confidence intervals for the mean of a normal distribution with known variance," *Annals of Math. Stat.*, Vol. 18 (1947), p. 427.
[23] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 434.
[24] A. Wald, "On the principles of statistical inference," *Notre Dame Math. Lectures*, No. 1 (1942).
[25] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Am. Math. Soc.*, Vol. 54 (1943), p. 426.
[26] A. Wald, "Statistical decision functions which minimize the maximum risk," *Annals of Math.*, Vol. 46 (1945), p. 265.
[27] A. Wald, "An essentially complete class of admissible decision functions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 549.
[28] A. Wald, *Sequential analysis*, John Wiley and Sons, 1947.
[29] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), p. 279.
[30] A. Wald, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), p. 165.
[31] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Math. Stat.*, Vol. 19 (1948), p. 326.
[32] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), p. 60.
[33] J. Wolfowitz, "Additive partition functions and a class of statistical hypotheses," *Annals of Math. Stat.*, Vol. 13 (1942), p. 247.
[34] J. Wolfowitz, "Non-parametric statistical inference," *Proceedings of the Berkeley Symposium on mathematical statistics and probability* (1949), p. 93.
[35] J. Wolfowitz, "The power of the classical tests associated with the normal distribution," *Annals of Math. Stat.*, Vol. 20 (1949), p. 540.
Some related papers not referred to in the text.
[36] T. W. Anderson, "On the theory of testing serial correlation," *Skand. Aktuarietidskrift*, (1948), p. 88.
[37] R. A. Fisher, *The design of experiments*, Oliver and Boyd, 1935.
[38] M. N. Ghosh, "On the problem of similar regions," *Sankhya*, Vol. 8 (1948), p. 329.
[39] P. G. Hoel, "Testing the homogeneity of Poisson frequencies," *Annals of Math. Stat.*, Vol. 16 (1945), p. 362.
[40] P. G. Hoel, "Discriminating between binomial distributions," *Annals of Math. Stat.*, Vol. 18 (1947), p. 556.
[41] P. G. Hoel, "On the uniqueness of similar regions," *Annals of Math. Stat.*, Vol. 19 (1948), p. 66.
[42] E. L. Lehmann, "Some comments on large sample tests," *Proceedings of the Berkeley Symposium on mathematical statistics and probability* (1949), p. 451.
[43] H. B. Mann and A. Wald, "On the choice of the number of intervals in the application of the χ^2-test," *Annals of Math. Stat.*, Vol. 13 (1942), p. 306.
[44] J. Neyman, *Lectures and conferences on mathematical statistics*, Graduate School of the U. S. Dept. of Agriculture, 1938.

[45] J. NEYMAN, "Basic ideas and some recent results of the theory of testing statistical hypotheses," *Journal Roy. Stat. Soc.*, Vol. cv (1942), p. 292.

[46] J. NEYMAN, "On a statistical problem arising in routine analysis and in sampling inspection of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), p. 46.

[47] S. N. ROY, "Notes on testing composite hypotheses, I, II," *Sankhya*, Vol. 8 (1947), p. 257 and Vol. 9 (1948), p. 19.

[48] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), p. 280.

[49] H. SCHEFFÉ, "On the ratio of the variances of two normal samples," *Annals of Math. Stat.*, Vol. 13 (1942), p. 371.

[50.] C. STEIN, "A two sample test for a linear hypothesis whose power is independent of the variance," *Annals of Math. Stat.*, Vol. 16 (1945), p. 243.

[51] A. WALD, "Asymptotically most powerful tests of statistical hypotheses," *Annals of* Math. Stat., Vol. 12 (1941), p. 1.

[52] A. WALD, "Some examples of asymptotically most powerful tests," *Annals of Math. Stat.*, Vol. 12 (1941), p. 396.

[53] A. WALD, "On the efficient design of statistical investigations," *Annals of Math. Stat.*, Vol. 14 (1943), p. 134.

SAMPLE CRITERIA FOR TESTING OUTLYING OBSERVATIONS[1]

BY FRANK E. GRUBBS

University of Michigan and Ballistic Research Laboratories

1. Summary. The problem of testing outlying observations, although an old one, is of considerable importance in applied statistics. Many and various types of significance tests have been proposed by statisticians interested in this field of application. In this connection, we bring out in the Historical Comments notable advances toward a clear formulation of the problem and important points which should be considered in attempting a complete solution. In Section 4 we state some of the situations the experimental statistician will very likely encounter in practice, these considerations being based on experience. For testing the significance of the largest observation in a sample of size n from a normal population, we propose the statistic

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $x_1 \leq x_2 \leq \cdots \leq x_n$, $\bar{x}_n = \dfrac{1}{n-1} \sum_{i=1}^{n-1} x_i$ *and* $\bar{x} = \dfrac{1}{n} \sum_{i=1}^{n} x_i$.

A similar statistic, S_1^2/S^2, can be used for testing whether the smallest observation is too low.

It turns out that

$$\frac{S_n^2}{S^2} = 1 - \frac{1}{n-1}\left(\frac{x_n - \bar{x}}{s}\right)^2 = 1 - \frac{1}{n-1} T_n^2,$$

where $s^2 = \dfrac{1}{n} \Sigma(x_i - \bar{x})^2$, and T_n is the studentized extreme deviation already suggested by E. Pearson and C. Chandra Sekar [1] for testing the significance of the largest observation. Based on previous work by W. R. Thompson [12], Pearson and Chandra Sekar were able to obtain certain percentage points of T_n without deriving the exact distribution of T_n. The exact distribution of S_n^2/S^2 (or T_n) is apparently derived for the first time by the present author.

For testing whether the two largest observations are too large we propose the statistic

$$\frac{S_{n-1,n}^2}{S^2} = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad \bar{x}_{n-1,n} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i$$

---

[1] This paper has been extracted from a thesis approved for the Degree of PhD at the University of Michigan.

and a similar statistic, $S_{1,2}^2/S^2$, can be used to test the significance of the two smallest observations. The probability distributions of the above sample statistics

TABLE I

Table of Percentage Points for $\frac{S_n^2}{S^2}$ *or* $\frac{S_1^2}{S^2}$

Percentage Points

n	1%	2.5%	5%	10%
3	.0001	.0007	.0027	.0109
4	.0100	.0248	.0494	.0975
5	.0442	.0808	.1270	.1984
6	.0928	.1453	.2032	.2826
7	.1447	.2066	.2696	.3503
8	.1948	.2616	.3261	.4050
9	.2411	.3101	.3742	.4502
10	.2831	.3526	.4154	.4881
11	.3211	.3901	.4511	.5204
12	.3554	.4232	.4822	.5483
13	.3864	.4528	.5097	.5727
14	.4145	.4792	.5340	.5942
15	.4401	.5030	.5559	.6134
16	.4634	.5246	.5755	.6306
17	.4848	.5442	.5933	.6461
18	.5044	.5621	.6095	.6601
19	.5225	.5785	.6243	.6730
20	.5393	.5937	.6379	.6848
21	.5548	.6076	.6504	.6958
22	.5692	.6206	.6621	.7058
23	.5827	.6327	.6728	.7151
24	.5953	.6439	.6829	.7238
25	.6071	.6544	.6923	.7319

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$S_n^2 = \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 \quad \text{where} \quad \bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i$$

$$S_1^2 = \sum_{i=2}^{n} (x_i - \bar{x}_1)^2 \quad \text{where} \quad \bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^{n} x_i$$

are derived for a normal parent and tables of appropriate percentage points are given in this paper (Table I and Table V). Although the efficiencies of the above tests have not been completely investigated under various models for outlying

observations, it is apparent that the proposed sample criteria have considerable intuitive appeal. In deriving the distributions of the sample statistics for testing the largest (or smallest) or the two largest (or two smallest) observations, it was first necessary to derive the distribution of the difference between the extreme observation and the sample mean in terms of the population σ. This probability

TABLE IA

Table of Percentage Points for $T_n = \dfrac{x_n - \bar{x}}{s}$ *or* $T_1 = \dfrac{\bar{x} - x_1}{s}$

n	1%	2.5%	5%	10%
3	1.414	1.414	1.412	1.406
4	1.723	1.710	1.689	1.645
5	1.955	1.917	1.869	1.791
6	2.130	2.067	1.996	1.894
7	2.265	2.182	2.093	1.974
8	2.374	2.273	2 172	2.041
9	2.464	2.349	2.237	2.097
10	2.540	2.414	2.294	2.146
11	2.606	2.470	2.343	2.190
12	2.663	2.519	2.387	2.229
13	2.714	2.562	2.426	2.264
14	2.759	2.602	2.461	2.297
15	2.800	2.638	2.493	2.326
16	2.837	2.670	2.523	2.354
17	2.871	2.701	2.551	2.380
18	2.903	2.728	2.577	2.404
19	2.932	2.754	2.600	2.426
20	2.959	2.778	2.623	2.447
21	2.984	2.801	2.644	2.467
22	3.008	2.823	2.664	2.486
23	3.030	2.843	2.683	2.504
24	3.051	2.862	2.701	2.520
25	3.071	2.880	2.717	2.537

$$x_1 \leq x_2 \leq x_3 \cdots \leq x_n$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

distribution was apparently derived first by A. T. McKay [11] who employed the method of characteristic functions. The author was not aware of the work of McKay when the simplified derivation for the distribution of $\dfrac{x_n - \bar{x}}{\sigma}$ outlined in Section 5 below was worked out by him in the spring of 1945, McKay's result

being called to his attention by C. C. Craig. It has been noted also that K. R. Nair [20] worked out independently and published the same derivation of the distribution of the extreme minus the mean arrived at by the present author—see *Biometrika*, Vol. 35, May, 1948. We nevertheless include part of this derivation in Section 5 below as it was basic to the work in connection with the derivations given in Sections 8 and 9. Our table is considerably more extensive than Nair's table of the probability integral of the extreme deviation from the sample mean in normal samples, since Nair's table runs from $n = 2$ to $n = 9$, whereas our Table II is for $n = 2$ to $n = 25$. The present work is concluded with some examples.

2. Introduction. Scientific data are collected usually for purposes of interpretation and if proper use is to be made of the information thus obtained then some decision should be reached or some action taken as a result of analyzing the data. In many cases a critical examination of the data collected is necessary in order to insure that the results of sampling are representative of the thing or process we are examining. Quite frequently our observations do not appear to be consistent with one another, i.e. the data may seem to display non-homogeneities and the group of observations as a whole may not appear to represent a random sample from, say, a single normal population or universe. In particular, one or more of the observations may have the appearance of being "outliers" and we are interested here in determining once and for all whether such observations should be retained in the sample for interpreting results or whether they should be regarded as being inconsistent with the remaining observations. It is clear that rejection of the "outliers" in a sample will in a great number of cases lead to a different course of action than would have been taken had such observations been retained in the sample. Actually, the rejection of "outlying" observations may be just as much a practical (or common sense) problem as a statistical one and sometimes the practical or experimental viewpoint may naturally outweigh any statistical contributions. In this connection, the concluding remarks of Rider's survey [2] are pertinent: "In the final analysis it would seem that the question of the rejection or the retention of a discordant observation reduces to a question of common sense. Certainly the judgment of an experienced observer should be allowed considerable influence in reaching a decision. This judgment can undoubtedly be aided by the application of one or more tests based on the theory of probability, but any test which requires an inordinate amount of calculation seems hardly to be worth while, and the testimony of any criterion which is based upon a complicated hypothesis should be accepted with extreme caution." Hence, it would appear that statistical tests of significance for judging or testing "outliers" come into importance either in supporting doubtful practical viewpoints or in providing a basis for action in the absence of sufficient experimental knowledge of underlying causes in an investigation. Indeed, the latter two situations are met quite frequently in practice.

In the present treatment, we intend to throw some light beyond the work

TABLE II

Probability Integral of the Extreme Minus the Mean, u_n, in Normal Samples of n Observations (Pop. S.D. as unit) $P(u_n \leq u)$

u \ n	2	3	4	5	6	7	8	9	u
.00	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.00
.05	.05637	.00309	.00017	.00001	.00000	.00000	.00000	.00000	.05
.10	.11246	.01231	.00134	.00015	.00002	.00000	.00000	.00000	.10
.15	.16800	.02745	.00445	.00072	.00012	.00002	.00000	.00000	.15
.20	.22270	.04817	.01033	.00221	.00047	.00010	.00002	.00000	.20
.25	.27633	.07403	.01966	.00520	.00137	.00036	.00010	.00003	.25
.30	.32863	.10450	.03292	.01033	.00324	.00101	.00032	.00010	.30
.35	.37938	.13896	.05040	.01820	.00656	.00236	.00085	.00031	.35
.40	.42839	.17677	.07218	.02935	.01191	.00482	.00195	.00079	.40
.45	.47548	.21724	.09816	.04416	.01982	.00889	.00398	.00178	.45
.50	.52050	.25968	.12807	.06288	.03080	.01507	.00737	.00360	.50
.55	.56332	.30344	.16152	.08559	.04525	.02390	.01261	.00665	.55
.60	.60386	.34788	.19801	.11219	.06344	.03583	.02022	.01140	.60
.65	.64203	.39243	.23697	.14246	.08547	.05121	.03067	.01836	.65
.70	.67780	.43656	.27781	.17602	.11130	.07030	.04437	.02800	.70
.75	.71116	.47983	.31992	.21242	.14076	.09318	.06164	.04076	.75
.80	.74210	.52185	.36274	.25113	.17353	.11978	.08263	.05698	.80
.85	.77067	.56230	.40571	.29160	.20920	.14993	.10739	.07688	.85
.90	.79691	.60095	.44835	.33325	.24727	.18329	.13578	.10055	.90
.95	.82089	.63761	.49021	.37555	.28721	.21945	.16757	.12791	.95
1.00	.84270	.67214	.53093	.41795	.32847	.25791	.20240	.15877	1.00
1.05	.86244	.70448	.57020	.45999	.37050	.29815	.23980	.19280	1.05
1.10	.88021	.73459	.60777	.50125	.41276	.33961	.27927	.22957	1.10
1.15	.89612	.76248	.64346	.54136	.45478	.38173	.32025	.26858	1.15
1.20	.91031	.78817	.67713	.58001	.49611	.42401	.36220	.30931	1.20
1.25	.92290	.81174	.70870	.61697	.53638	.46595	.40457	.35117	1.25
1.30	.93401	.83325	.73812	.65205	.57525	.50712	.44685	.39362	1.30
1.35	.94376	.85280	.76540	.68513	.61249	.54716	.48857	.43613	1.35
1.40	.95229	.87049	.79055	.71612	.64788	.58574	.52933	.47822	1.40
1.45	.95970	.88644	.81364	.74497	.68129	.62263	.56878	.51945	1.45
1.50	.96611	.90075	.83472	.77170	.71261	.65762	.60663	.55944	1.50
1.55	.97162	.91355	.85390	.79632	.74180	.69058	.64265	.59789	1.55
1.60	.97635	.92495	.87127	.81890	.76885	.72143	.67668	.63456	1.60
1.65	.98038	.93506	.88693	.83949	.79378	.75013	.70862	.66925	1.65
1.70	.98379	.94400	.90099	.85820	.81664	.77666	.73839	.70184	1.70
1.75	.98667	.95187	.91358	.87513	.83750	.80107	.76597	.73225	1.75
1.80	.98909	.95877	.92480	.89037	.85646	.82341	.79139	.76046	1.80
1.85	.99111	.96480	.93476	.90405	.87360	.84376	.81469	.78647	1.85
1.90	.99279	.97005	.94358	.91628	.88903	.86220	.83593	.81032	1.90
1.95	.99418	.97461	.95135	.92716	.90288	.87885	.85522	.83207	1.95

TABLE II—*Continued*

u \ n	2	3	4	5	6	7	8	9	u
2.00	.99532	.97854	.95818	.93682	.91526	.89381	.87264	.85183	2.00
2.05	.99626	.98193	.96416	.94536	.92627	.90721	.88832	.86968	2.05
2.10	.99702	.98483	.96938	.95289	.93605	.91916	.90236	.88574	2.10
2.15	.99764	.98731	.97392	.95949	.94468	.92977	.91490	.90012	2.15
2.20	.99814	.98942	.97785	.96527	.95229	.93917	.92604	.91296	2.20
2.25	.99854	.99121	.98125	.97032	.95897	.94746	.93591	.92438	2.25
2.30	.99886	.99273	.98418	.97470	.96482	.95476	.94462	.93448	2.30
2.35	.99911	.99400	.98669	.97850	.96992	.96114	.95229	.94340	2.35
2.40	.99931	.99507	.98883	.98178	.97435	.96672	.95900	.95125	2.40
2.45	.99947	.99596	.99066	.98461	.97819	.97158	.96487	.95812	2.45
2.50	.99959	.99670	.99222	.98703	.98151	.97580	.96999	.96412	2.50
2.55	.99969	.99732	.99353	.98911	.98436	.97944	.97443	.96935	2.55
2.60	.99976	.99782	.99464	.99088	.98681	.98259	.97827	.97389	2.60
2.65	.99982	.99824	.99557	.99238	.98891	.98529	.98158	.97781	2.65
2.70	.99987	.99858	.99635	.99365	.99070	.98761	.98443	.98120	2.70
2.75	.99990	.99886	.99701	.99473	.99223	.98959	.98688	.98411	2.75
2.80	.99992	.99909	.99755	.99564	.99352	.99128	.98897	.98661	2.80
2.85	.99994	.99928	.99800	.99640	.99461	.99272	.99075	.98874	2.85
2.90	.99996	.99943	.99838	.99704	.99553	.99393	.99227	.99056	2.90
2.95	.99997	.99955	.99868	.99757	.99631	.99496	.99355	.99211	2.95
3.00	.99998	.99964	.99894	.99801	.99696	.99582	.99464	.99342	3.00
3.05	.99998	.99972	.99914	.99838	.99750	.99655	.99555	.99453	3.05
3.10	.99999	.99978	.99931	.99868	.99795	.99716	.99632	.99546	3.10
3.15	.99999	.99983	.99945	.99893	.99832	.99766	.99697	.99625	3.15
3.20	.99999	.99987	.99956	.99913	.99863	.99808	.99750	.99690	3.20
3.25	1.00000	.99990	.99965	.99930	.99889	.99843	.99795	.99745	3.25
3.30		.99992	.99972	.99944	.99910	.99872	.99832	.99791	3.30
3.35		.99994	.99978	.99955	.99927	.99896	.99863	.99829	3.35
3.40		.99995	.99983	.99964	.99941	.99916	.99889	.99860	3.40
3.45		.99996	.99986	.99971	.99953	.99932	.99910	.99886	3.45
3.50		.99997	.99989	.99977	.99962	.99945	.99927	.99908	3.50
3.55		.99998	.99992	.99982	.99970	.99956	.99941	.99925	3.55
3.60		.99998	.99994	.99986	.99976	.99965	.99952	.99940	3.60
3.65		.99999	.99995	.99989	.99981	.99972	.99962	.99951	3.65
3.70		.99999	.99996	.99991	.99985	.99977	.99969	.99961	3.70
3.75		.99999	.99997	.99993	.99988	.99982	.99976	.99969	3.75
3.80		1.00000	.99998	.99995	.99991	.99986	.99981	.99975	3.80
3.85			.99998	.99996	.99993	.99989	.99985	.99980	3.85
3.90			.99999	.99997	.99994	.99991	.99988	.99984	3.90
3.95			.99999	.99997	.99995	.99993	.99990	.99987	3.95

TABLE II—*Continued*

u \ n	2	3	4	5	6	7	8	9	u
4.00			.99999	.99998	.99996	.99995	.99992	.99990	4.00
4.05			.99999	.99999	.99997	.99996	.99994	.99992	4.05
4.10			1.00000	.99999	.99998	.99997	.99995	.99994	4.10
4.15				.99999	.99998	.99997	.99996	.99995	4.15
4.20				.99999	.99999	.99998	.99997	.99996	4.20
4.25				.99999	.99999	.99998	.99998	.99997	4.25
4.30				1.00000	.99999	.99999	.99998	.99998	4.30
4.35					.99999	.99999	.99999	.99998	4.35
4.40					1.00000	.99999	.99999	.99999	4.40
4.45						.99999	.99999	.99999	4.45
4.50						1.00000	.99999	.99999	4.50
4.55							1.00000	.99999	4.55
4.60								1.00000	4.60

u \ n	10	11	12	13	14	15	16	17	u
.25	.00001	.00000	.00000	.00000	.00000	.00000	.00000	.00000	.25
.30	.00003	.00001	.00000	.00000	.00000	.00000	.00000	.00000	.30
.35	.00011	.00004	.00001	.00001	.00000	.00000	.00000	.00000	.35
.40	.00032	.00013	.00005	.00002	.00001	.00000	.00000	.00000	.40
.45	.00080	.00036	.00016	.00007	.00003	.00001	.00001	.00000	.45
.50	.00176	.00086	.00042	.00021	.00010	.00005	.00002	.00001	.50
.55	.00351	.00185	.00098	.00051	.00027	.00014	.00008	.00004	.55
.60	.00643	.00363	.00204	.00115	.00065	.00037	.00021	.00012	.60
.65	.01098	.00657	.00393	.00235	.00141	.00084	.00050	.00030	.65
.70	.01766	.01113	.00702	.00443	.00279	.00176	.00111	.00070	.70
.75	.02694	.01780	.01177	.00777	.00514	.00339	.00224	.00148	.75
.80	.03928	.02707	.01865	.01285	.00886	.00610	.00420	.00289	.80
.85	.05503	.03938	.02818	.02016	.01442	.01031	.00738	.00527	.85
.90	.07444	.05510	.04077	.03017	.02232	.01652	.01222	.00904	.90
.95	.09761	.07448	.05682	.04334	.03305	.02521	.01922	.01466	.95
1.00	.12452	.09763	.07655	.06000	.04703	.03687	.02889	.02265	1.00
1.05	.15497	.12454	.10008	.08041	.06460	.05190	.04169	.03348	1.05
1.10	.18867	.15503	.12737	.10464	.08595	.07060	.05799	.04762	1.10
1.15	.22520	.18879	.15825	.13263	.11116	.09315	.07806	.06541	1.15
1.20	.26407	.22542	.19240	.16420	.14013	.11957	.10203	.08706	1.20
1.25	.30475	.26442	.22941	.19901	.17263	.14973	.12987	.11264	1.25
1.30	.34666	.30525	.26876	.23662	.20830	.18336	.16140	.14207	1.30
1.35	.38924	.34734	.30992	.27650	.24667	.22005	.19629	.17509	1.35
1.40	.43196	.39011	.35229	.31810	.28721	.25931	.23411	.21135	1.40
1.45	.47430	.43302	.39529	.36082	.32934	.30058	.27433	.25036	1.45

TABLE II—*Continued*

u \ n	10	11	12	13	14	15	16	17	u
1.50	.51583	.47555	.43838	.40408	.37244	.34327	.31636	.29156	1.50
1.55	.55615	.51726	.48104	.44733	.41595	.38676	.35960	.33434	1.55
1.60	.59495	.55774	.52282	.49004	.45930	.43046	.40342	.37807	1.60
1.65	.63196	.59668	.56332	.53178	.50199	.47384	.44726	.42216	1.65
1.70	.66699	.63380	.60221	.57216	.54358	.51641	.49058	.46602	1.70
1.75	.69991	.66892	.63925	.61086	.58370	.55773	.53289	.50915	1.75
1.80	.73063	.70189	.67424	.64763	.62204	.59744	.57380	.55108	1.80
1.85	.75912	.73264	.70704	.68229	.65838	.63528	.61297	.59144	1.85
1.90	.78538	.76113	.73758	.71472	.69254	.67102	.65016	.62992	1.90
1.95	.80945	.78737	.76584	.74486	.72443	.70453	.68516	.66630	1.95
2.00	.83141	.81140	.79183	.77269	.75399	.73571	.71786	.70042	2.00
2.05	.85133	.83330	.81560	.79824	.78121	.76453	.74819	.73218	2.05
2.10	.86932	.85314	.83721	.82155	.80614	.79101	.77614	.76153	2.10
2.15	.88550	.87105	.85678	.84271	.82885	.81519	.80174	.78849	2.15
2.20	.89998	.88713	.87440	.86183	.84941	.83715	.82505	.81311	2.20
2.25	.91290	.90151	.89021	.87902	.86795	.85699	.84616	.83545	2.25
2.30	.92437	.91431	.90432	.89441	.88458	.87484	.86518	.85563	2.30
2.35	.93453	.92568	.91688	.90812	.89943	.89081	.88224	.87375	2.35
2.40	.94348	.93572	.92799	.92030	.91264	.90504	.89748	.88997	2.40
2.45	.95134	.94457	.93781	.93106	.92435	.91766	.91101	.90440	2.45
2.50	.95823	.95233	.94644	.94055	.93468	.92883	.92300	.91720	2.50
2.55	.96424	.95912	.95400	.94887	.94376	.93866	.93357	.92850	2.55
2.60	.96948	.96504	.96060	.95616	.95172	.94728	.94285	.93844	2.60
2.65	.97401	.97019	.96635	.96251	.95866	.95482	.95098	.94715	2.65
2.70	.97793	.97464	.97134	.96802	.96471	.96139	.95807	.95475	2.70
2.75	.98131	.97849	.97565	.97280	.96995	.96709	.96423	.96137	2.75
2.80	.98422	.98180	.97937	.97693	.97448	.97203	.96957	.96712	2.80
2.85	.98671	.98464	.98257	.98048	.97839	.97629	.97418	.97208	2.85
2.90	.98883	.98708	.98531	.98353	.98174	.97995	.97816	.97636	2.90
2.95	.99064	.98915	.98765	.98614	.98462	.98309	.98156	.98003	2.95
3.00	.99218	.99092	.98965	.98837	.98708	.98578	.98448	.98318	3.00
3.05	.99348	.99242	.99134	.99026	.98917	.98807	.98697	.98587	3.05
3.10	.99458	.99369	.99278	.99187	.99095	.99002	.98909	.98816	3.10
3.15	.99551	.99476	.99400	.99323	.99245	.99167	.99089	.99010	3.15
3.20	.99628	.99566	.99502	.99437	.99372	.99307	.99241	.99175	3.20
3.25	.99694	.99641	.99588	.99534	.99479	.99424	.99369	.99314	3.25
3.30	.99748	.99704	.99660	.99615	.99569	.99523	.99477	.99431	3.30
3.35	.99793	.99757	.99720	.99682	.99644	.99606	.99568	.99529	3.35
3.40	.99831	.99801	.99770	.99739	.99707	.99676	.99644	.99611	3.40
3.45	.99862	.99837	.99812	.99786	.99760	.99733	.99707	.99680	3.45

TABLE II—*Continued*

u \ n	10	11	12	13	14	15	16	17	u
3.50	.99888	.99867	.99846	.99825	.99803	.99781	.99759	.99737	3.50
3.55	.99909	.99892	.99875	.99857	.99839	.99821	.99803	.99785	3.55
3.60	.99926	.99912	.99898	.99884	.99869	.99854	.99839	.99824	3.60
3.65	.99940	.99929	.99917	.99906	.99894	.99881	.99869	.99857	3.65
3.70	.99952	.99943	.99933	.99924	.99914	.99904	.99894	.99883	3.70
3.75	.99961	.99954	.99946	.99938	.99930	.99922	.99914	.99905	3.75
3.80	.99969	.99963	.99957	.99950	.99944	.99937	.99930	.99923	3.80
3.85	.99975	.99970	.99965	.99960	.99955	.99949	.99944	.99938	3.85
3.90	.99980	.99976	.99972	.99968	.99964	.99959	.99955	.99950	3.90
3.95	.99984	.99981	.99978	.99974	.99971	.99967	.99964	.99960	3.95
4.00	.99988	.99985	.99982	.99980	.99977	.99974	.99971	.99968	4.00
4.05	.99990	.99988	.99986	.99984	.99982	.99979	.99977	.99974	4.05
4.10	.99992	.99991	.99989	.99987	.99985	.99983	.99981	.99979	4.10
4.15	.99994	.99993	.99991	.99990	.99988	.99987	.99985	.99984	4.15
4.20	.99995	.99994	.99993	.99992	.99991	.99990	.99988	.99987	4.20
4.25	.99996	.99995	.99995	.99994	.99993	.99992	.99991	.99990	4.25
4.30	.99997	.99996	.99996	.99995	.99994	.99993	.99993	.99992	4.30
4.35	.99998	.99997	.99997	.99996	.99996	.99995	.99994	.99993	4.35
4.40	.99998	.99998	.99997	.99997	.99996	.99996	.99995	.99995	4.40
4.45	.99999	.99998	.99998	.99998	.99997	.99997	.99996	.99996	4.45
4.50	.99999	.99999	.99998	.99998	.99998	.99998	.99997	.99997	4.50
4.55	.99999	.99999	.99999	.99999	.99998	.99998	.99998	.99997	4.55
4.60	.99999	.99999	.99999	.99999	.99999	.99998	.99998	.99998	4.60
4.65	1.00000	.99999	.99999	.99999	.99999	.99999	.99999	.99998	4.65
4.70		1.00000	.99999	.99999	.99999	.99999	.99999	.99999	4.70
4.75			1.00000	1.00000	.99999	.99999	.99999	.99999	4.75
4.80					1.00000	.99999	.99999	.99999	4.80
4.85						1.00000	1.00000	1.00000	4.85

u \ n	18	19	20	21	22	23	24	25	u
.50	.00001	.00000	.0000	.0000	.0000	.0000	.0000	.0000	.50
.55	.00002	.00001	.0000	.0000	.0000	.0000	.0000	.0000	.55
.60	.00007	.00004	.0000	.0000	.0000	.0000	.0000	.0000	.60
.65	.00018	.00011	.0001	.0000	.0000	.0000	.0000	.0000	.65
.70	.00044	.00028	.0002	.0001	.0001	.0000	.0000	.0000	.70
.75	.00098	.00065	.0004	.0003	.0002	.0001	.0001	.0001	.75
.80	.00199	.00137	.0009	.0007	.0004	.0003	.0002	.0001	.80
.85	.00377	.00270	.0019	.0014	.0010	.0007	.0005	.0004	.85
.90	.00669	.00494	.0037	.0027	.0020	.0015	.0011	.0008	.90
.95	.01118	.00853	.0065	.0049	.0038	.0029	.0022	.0017	.95

TABLE II—*Continued*

u \ n	18	19	20	21	22	23	24	25	u
1.00	.01775	.01391	.0109	.0085	.0067	.0052	.0041	.0032	1.00
1.05	.02690	.02161	.0174	.0139	.0112	.0090	.0072	.0058	1.05
1.10	.03911	.03212	.0264	.0217	.0178	.0146	.0120	.0099	1.10
1.15	.05481	.04592	.0385	.0322	.0270	.0226	.0190	.0159	1.15
1.20	.07428	.06338	.0541	.0461	.0394	.0336	.0287	.0244	1.20
1.25	.09769	.08472	.0735	.0637	.0553	.0479	.0416	.0360	1.25
1.30	.12504	.11005	.0969	.0853	.0750	.0660	.0581	.0512	1.30
1.35	.15618	.13930	.1242	.1108	.0988	.0882	.0786	.0701	1.35
1.40	.19080	.17225	.1555	.1404	.1267	.1144	.1033	.0932	1.40
1.45	.22848	.20851	.1903	.1736	.1585	.1446	.1320	.1204	1.45
1.50	.26869	.24761	.2282	.2103	.1938	.1786	.1646	.1516	1.50
1.55	.31084	.28899	.2687	.2498	.2322	.2159	.2007	.1866	1.55
1.60	.35430	.33202	.3111	.2916	.2732	.2560	.2399	.2248	1.60
1.65	.39845	.37607	.3549	.3349	.3162	.2984	.2816	.2658	1.65
1.70	.44269	.42052	.3994	.3794	.3604	.3424	.3252	.3089	1.70
1.75	.48645	.46476	.4440	.4242	.4053	.3872	.3699	.3534	1.75
1.80	.52924	.50827	.4881	.4687	.4502	.4323	.4152	.3987	1.80
1.85	.57065	.55058	.5312	.5125	.4945	.4771	.4603	.4441	1.85
1.90	.61031	.59130	.5729	.5549	.5377	.5209	.5047	.4890	1.90
1.95	.64796	.63011	.6127	.5958	.5794	.5634	.5479	.5328	1.95
2.00	.68340	.66678	.6506	.6348	.6193	.6042	.5895	.5752	2.00
2.05	.71650	.70114	.6861	.6714	.6570	.6429	.6291	.6156	2.05
2.10	.74719	.73311	.7193	.7058	.6924	.6793	.6665	.6540	2.10
2.15	.77545	.76262	.7500	.7375	.7254	.7133	.7015	.6899	2.15
2.20	.80132	.78971	.7782	.7670	.7558	.7448	.7340	.7234	2.20
2.25	.82486	.81440	.8041	.7938	.7838	.7738	.7640	.7543	2.25
2.30	.84616	.83679	.8275	.8184	.8093	.8003	.7914	.7827	2.30
2.35	.86533	.85699	.8487	.8405	.8324	.8244	.8164	.8085	2.35
2.40	.88251	.87511	.8678	.8605	.8533	.8461	.8390	.8319	2.40
2.45	.89783	.89129	.8848	.8784	.8720	.8656	.8593	.8530	2.45
2.50	.91142	.90568	.9000	.8943	.8887	.8831	.8775	.8719	2.50
2.55	.92345	.91842	.9134	.9084	.9035	.8985	.8936	.8888	2.55
2.60	.93404	.92965	.9253	.9209	.9166	.9123	.9080	.9037	2.60
2.65	.94332	.93951	.9357	.9319	.9282	.9244	.9207	.9169	2.65
2.70	.95144	.94814	.9448	.9416	.9382	.9351	.9318	.9286	2.70
2.75	.95852	.95567	.9528	.9500	.9472	.9444	.9415	.9387	2.75
2.80	.96466	.96220	.9598	.9573	.9549	.9524	.9500	.9476	2.80
2.85	.96997	.96787	.9658	.9637	.9616	.9595	.9574	.9553	2.85
2.90	.97456	.97275	.9710	.9692	.9674	.9656	.9638	.9620	2.90
2.95	.97850	.97696	.9754	.9739	.9724	.9709	.9693	.9678	2.95

TABLE II—*Continued*

u \ n	18	19	20	21	22	23	24	25	u
3.00	.98187	.98057	.9793	.9780	.9767	.9753	.9741	.9728	3.00
3.05	.98476	.98365	.9825	.9814	.9803	.9793	.9781	.9771	3.05
3.10	.98722	.98629	.9853	.9844	.9835	.9826	.9816	.9807	3.10
3.15	.98931	.98852	.9877	.9869	.9862	.9853	.9846	.9838	3.15
3.20	.99108	.99042	.9898	.9891	.9884	.9878	.9871	.9865	3.20
3.25	.99258	.99202	.9915	.9909	.9904	.9898	.9893	.9887	3.25
3.30	.99384	.99337	.9929	.9924	.9920	.9915	.9911	.9906	3.30
3.35	.99490	.99451	.9941	.9937	.9933	.9930	.9926	.9922	3.35
3.40	.99579	.99546	.9951	.9948	.9945	.9942	.9939	.9936	3.40
3.45	.99653	.99626	.9960	.9957	.9955	.9952	.9949	.9947	3.45
3.50	.99715	.99693	.9967	.9965	.9963	.9961	.9958	.9956	3.50
3.55	.99766	.99748	.9973	.9971	.9969	.9968	.9966	.9964	3.55
3.60	.99809	.99794	.9978	.9976	.9975	.9973	.9972	.9971	3.60
3.65	.99844	.99832	.9982	.9981	.9979	.9978	.9977	.9976	3.65
3.70	.99873	.99863	.9985	.9984	.9983	.9982	.9982	.9981	3.70
3.75	.99897	.99889	.9988	.9987	.9986	.9986	.9985	.9984	3.75
3.80	.99917	.99910	.9990	.9990	.9989	.9988	.9988	.9988	3.80
3.85	.99933	.99927	.9992	.9992	.9991	.9991	.9990	.9990	3.85
3.90	.99946	.99941	.9994	.9993	.9993	.9993	.9992	.9992	3.90
3.95	.99956	.99953	.9995	.9995	.9994	.9994	.9994	.9994	3.95
4.00	.99965	.99962	.9996	.9996	.9995	.9995	.9995	.9995	4.00
4.05	.99972	.99969	.9997	.9996	.9996	.9996	.9996	.9996	4.05
4.10	.99977	.99975	.9997	.9997	.9997	.9997	.9997	.9997	4.10
4.15	.99982	.99980	.9998	.9998	.9998	.9998	.9998	.9998	4.15
4.20	.99986	.99984	.9998	.9998	.9998	.9998	.9998	.9998	4.20
4.25	.99989	.99987	.9999	.9999	.9999	.9999	.9999	.9999	4.25
4.30	.99991	.99990	.9999	.9999	.9999	.9999	.9999	.9999	4.30
4.35	.99993	.99992	.9999	.9999	.9999	.9999	.9999	.9999	4.35
4.40	.99994	.99994	.9999	.9999	.9999	.9999	.9999	.9999	4.40
4.45	.99995	.99995	1.0000	.9999	.9999	.9999	.9999	.9999	4.45
4.50	.99996	.99996		1.0000	1.0000	1.0000	.9999	.9999	4.50
4.55	.99997	.99997					1.0000	1.0000	4.55
4.60	.99998	.99997							4.60
4.65	.99998	.99998							4.65
4.70	.99998	.99998							4.70
4.75	.99999	.99998							4.75
4.80	.99999	.99999							4.80
4.85	.99999	.99999							4.85
4.90	1.00000	1.00000							4.90

that has already been done [1], [2], [3], [4], [11], [12], [20] on the problem of testing outlying observations statistically and to see just where our contributions fit into this corner of mathematical statistics. First, however, we give a very brief history of the problem.

3. Historical comments. A survey of statistical literature indicates that the problem of testing the significance of outlying observations received considerable attention prior to 1937. Since this date, however, published literature on the subject seems to have been unusually scant—perhaps because of inherent difficulties in the problem as pointed out by E. S. Pearson and C. Chandra Sekar [1]. These authors made some important contributions to the problem of outlying observations by bringing clearly into the foreground the concept of efficiency of tests which may be used in view of admissible alternative hypotheses.

In 1933, P. R. Rider [2] published a rather comprehensive survey of work on the problem of testing the significance of outlying observations up to that date. The test criteria surveyed by Rider appear to impose as an initial condition that the standard deviation, σ, of the population from which the items were drawn should be known accurately. In connection with such tests requiring accurate knowledge of σ, we mention (1) Irwin's criteria [3] which utilize the difference between the first two individuals or the difference between the second and third individuals in random samples from a normal population and (2) the range[2] or maximum dispersion [4], [5], [6], [7], [8], [9], [10], [18] of a sample which has been advocated by "Student" [4] and others for testing the significance of outlying observations. We remark further that a natural statistic to use for testing an "outlier" is the difference between such an extreme observation and the sample mean. In 1935, McKay [11] published a note on the distribution of the last-mentioned statistic and by means of a rather elaborate procedure obtained a recurrence relation between the distribution of the extreme minus the mean in samples of n from a normal universe and the distribution of this statistic in samples of $n - 1$ from the same parent. McKay gave also an approximate expression for the upper percentage points of the distribution but did not tabulate the exact distribution due to the complicity of the multiple integrals involved. McKay pointed out that if K_p denotes the p-th semi-invariant of the distribution of $x_n - \bar{x}$ (where x_n is the largest observation) and K'_p refers similarly to the distribution of x_n, then $K_1 = K'_1 - \mu$, $K_2 = K'_2 - \frac{1}{n}$ and $K_p = K'_p$ ($p \geq 3$ where $\mu = E(x_i)$. Nair [20] has tabulated the distribution of the difference between the extreme and sample mean for $n = 2$ to $n = 9$.

Under certain circumstances, accurate knowledge concerning σ may be available as, for example, in using "daily control" tests [4], [18] the population standard deviation may be estimated in some cases with sufficient precision from past

[2] The derivation for the exact distribution of the range is given in reference [9], 1942; however, Dr. L. S. Dederick of the Ballistic Research Laboratory also derived the exact distribution of the range in an unpublished Aberdeen Proving Ground Report (1926).

data. In general, however, an accurate estimate of σ may not be available and it becomes necessary to estimate the population standard deviation from the single sample involved or "Studentize" [18], [20] the statistic to be used, thus providing a true measure of the risks involved in the significance test advocated for testing outlying observations. W. R. Thompson [12] apparently had this very point in mind when he devised an exact test in his paper, "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation," which appeared in 1935. Thompson showed that if

$$T_i = \frac{x_i - \bar{x}}{s}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and x_i is an observation selected arbitrarily from a random sample of n items drawn from a normal parent, then the probability density function of

$$t = \frac{T\sqrt{n-2}}{\sqrt{n-1-T^2}}$$

is given by "Student's" t-distribution with $f = n - 2$ degrees of freedom.

Pearson and Chandra Sekar have given a rather comprehensive study of Thompson's criterion in an interesting and important paper [1] which appeared in 1936. They discussed also some very important viewpoints which should be taken into consideration when dealing with the problem of testing outlying observations. By setting up alternatives to the null-hypothesis H_0 that all items in the sample come from the same population, Pearson and Chandra Sekar point out that if only one of the observations actually came from a population with divergent mean, then Thompson's criterion would be very useful, whereas if two or more of the observations are truly outlying then the criterion $| x_i - \bar{x} | \geq T_0 s$ may be quite ineffective, particularly if the sample contains less than about 30 or 40 observations.

A point of major interest concerning Thompson's work nevertheless is that he proposed an *exact* test for the hypothesis that all of the observations came from the same normal population. With regard to the use of an arbitrary observation in Thompson's test, however, it should be borne in mind that the problem of finding the probability that an arbitrary observation will be outlying is different from that of finding the probability that a particular observation (the largest, for example) will be outlying with respect to the other $n - 1$ observations of the sample.

As a final point concerning the paper of Pearson and Chandra Sekar [1], we see that for the n values of T_i arranged in order of magnitude taking account of sign, say

$$T^{(1)}, T^{(2)}, \cdots T^{(n)},$$

then

$$T^{(1)} \geq T^{(2)} \geq T^{(3)} \cdots \geq T^{(n)}.$$

The above authors show that the form of the total distribution of all the T_i at its extremes depend only on $T^{(1)}$ and $T^{(n)}$. This is because for some combinations of sample size and percentage points the algebraic upper limit for $T^{(2)}$ and algebraic lower limit for $T^{(n-1)}$ do not extend into the "tails" of the total distribution. Hence, the following probability law holds for $T^{(1)}$ when $T^{(1)} \geq$ the algebraic maximum of $T^{(2)}$:

$$p\{T^{(1)}\} = Np(T).$$

Likewise,

$$p\{T^{(n)}\} = Np(T)$$

for $T^{(n)} \leq$ algebraic minimum of $T^{(n-1)}$. Therefore, Pearson and Chandra Sekar were able to use Thompson's table [12] and give (for some sample sizes) upper probability limits for $T^{(1)} = \dfrac{x_i - \bar{x}}{s}$ for the highest observation and lower probability limits for $T^{(n)} = \dfrac{x_i - \bar{x}}{s}$ for the lowest observation without actually obtaining the exact probability distribution of $T^{(1)}$ and $T^{(n)}$. Hence, the appearance of the table of percentage points on page 318 of their paper [1] was a substantial contribution to the problem of testing outlying observations since an exact test for the significance of a single outlying observation was provided for the case where an accurate estimate of σ is not available. (The exact distribution of $T^{(1)}$ or $T^{(n)}$ is derived later in this work.)

With the above highlights of historical background in mind, we turn now to a consideration of the types of problems the experimenter may be faced with in testing "outlying" observations.

4. Statement of hypotheses in tests of outliers. Once the sample results of an experiment are available, the practicing statistician may be confronted with one or more of the following distinct situations as regards discordant observations: (a) To begin with, a very frequent or perhaps prevalent situation is that either the greatest observation or the least observation in a sample may have the appearance of belonging to a different population than the one from which the remaining observations were drawn. Here we are confronted with tests for a single outlying observation. (b) Then again, both the largest and the smallest observations may appear to be "different" from the remaining items in the sample. Here we are interested in testing the hypothesis that both the largest and the smallest observations are truly "outliers." (c) Another frequent situation is that either the two largest or the two smallest observations may have the appearance of being discordant. Here we are interested in reaching a decision as to whether we should reject the two largest or the two smallest observations as not being representative of the thing we are sampling.

As to why the discordant observations in a sample may be outliers, this may be due to errors of measurement in which case we would naturally want to reject or at least "correct" such observations. On the other hand, it may be that the population we are sampling is not homogeneous in the uni-modal sense and it will consequently be desirable to know this so that we may carry out further development work on our product if possible or desirable.

Although there may be many models for outliers, we believe that an important practical case involves the situation where all the observations in the sample may be subject to the same standard error, whereas it may happen that the largest or smallest observations result from shifts in level. For example, if one observation appears unusually high compared to the others in the sample we may want to consider the hypothesis that all the observations come from a normal parent with mean μ and standard deviation σ as against the alternative hypothesis that the largest observation comes from a normal population with mean $\mu + \lambda\sigma$ $(\lambda > 0)$ and standard deviation σ, whereas the remaining observations are from $N(\mu, \sigma)$.

Another case involves the situation where the largest and/or smallest observations may be from $N(\mu, \lambda\sigma)$, $\lambda > 1$, whereas the remaining observations of the sample are from the normal parent $N(\mu, \sigma)$.

Although we have not investigated the power of the tests proposed herein for various models, it is believed that the exact test of Section 8 for the largest (or smallest) observation and the test of Section 9 for the two largest (or two smallest) observations possess considerable intuitive appeal for the practical situations described above.[3]

5. Distribution of the difference between the extreme and mean in samples of n from a normal population. The simultaneous density function of n independent observations from a normal parent with zero mean and variance σ^2 which are arranged in order of magnitude is given by

$$dF(x_1, x_2, \cdots, x_n) = \frac{n!}{(\sqrt{2\pi}\,\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right] \quad dx_1\, dx_2 \cdots dx_n \tag{1}$$

subject to $x_1 \leq x_2 \leq \cdots \leq x_n$.

Since

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n}{n-1}(x_n - \bar{x})^2 + \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2$$

where

$$\bar{x}_n = \frac{1}{n-1}\sum_{i=1}^{n-1} x_i,$$

[3] The author is indebted to J. W. Tukey and S. S. Wilks for calling attention to an incorrect distribution function in the originally submitted manuscript on which several yet-to-be proved or disproved statements concerning optimum properties of statistics in this paper were based.

then

$$\sum_{i=1}^{n} x_i^2 = n\bar{x}^2 + \frac{n}{n-1}(x_n - \bar{x})^2 + \frac{n-1}{n-2}(x_{n-1} - \bar{x}_n)^2$$

$$(2) \qquad + \frac{n-2}{n-3}(x_{n-2} - \bar{x}_{n,n-1})^2 + \cdots + \frac{3}{2}\left(x_3 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \frac{2}{1}\left(x_2 - \frac{x_1 + x_2}{2}\right)^2$$

where

$$\bar{x}_{n,n-1} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i, \quad \text{etc.}$$

and consequently we find that we are particularly interested in the following Helmert orthogonal transformation:

$$\sqrt{2\cdot 1}\,\sigma\eta_2 = -x_1 + x_2,$$

$$\sqrt{3\cdot 2}\,\sigma\eta_3 = -x_1 - x_2 + 2x_3,$$

$$(3) \qquad \vdots$$

$$\sqrt{n(n-1)}\,\sigma\eta_n = -x_1 - x_2 - x_3 - x_4 - \cdots - x_r - \cdots - x_{n-1} + (n-1)x_n,$$

$$\sqrt{n}\,\sigma\eta_{n+1} = x_1 + x_2 + x_3 + x_4 + \cdots + x_r + \cdots + x_{n-1} + x_n.$$

The above transformation will lead to the distribution of the difference between the extreme and sample mean in terms of the unknown population σ for samples of n from a normal parent. Since, however, K. R. Nair (*Biometrika*, May, 1948) has already published the details independently, we will only record here for later reference that the density function of $\eta_2, \eta_3, \cdots, \eta_n$ (after integrating η_{n+1} over $-\infty \leq \eta_{n+1} \leq +\infty$) is

$$(4) \qquad dF(\eta_2, \eta_3, \cdots \eta_n) = \frac{n!}{(\sqrt{2\pi})^{n-1}} \exp\left[-\frac{1}{2}\sum_{i=2}^{n} \eta_i^2\right] d\eta_2\, d\eta_3 \cdots d\eta_n$$

where the η_i are restricted by the relations

$$(5) \qquad \infty \geq \eta_2 \geq 0, \qquad \sqrt{\frac{r}{r-2}}\,\eta_r \geq \eta_{r-1}.$$

Upon making the transformations

$$(6) \qquad \frac{\sqrt{r(r-1)}}{r}\,\eta_r = \frac{x_r - \bar{x}}{\sigma} = u_r, \qquad (r = 2, 3, \cdots, n),$$

defining

$$(7) \qquad F_n(u) = \int_0^u dF(u_n) = \text{probability } u_n \leq u,$$

and integrating the u_n over their appropriate ranges we find the cumulative probability integrals of the extreme deviation from the sample mean (in terms of the population σ) for $n = 2, 3, \cdots$ to be

$$F_2(u) = 2\sqrt{2}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2x^2)}\, dx = \frac{2}{\sqrt{\pi}}\int_0^u e^{-x^2}\, dx,$$

a well-known result, where for $n = 2$, x is either the sample standard deviation, the difference between the extreme and sample mean, the mean deviation or the semi-range.

$$F_3(u) = \frac{3\sqrt{3}}{\sqrt{2}}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{3}{2}x^2)}\, F_2(\tfrac{3}{2}x)\, dx\,,$$

(8)
$$\vdots$$

$$F_n(u) = \frac{n\sqrt{n}}{\sqrt{n-1}}\int_0^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}((n)/(n-1)x^2)}\, F_{n-1}\left(\frac{n}{n-1}\,x\right) dx\,.$$

This is equivalent to the result of McKay (11), although the derivation indicated is a considerably simpler one.

Now $F_{n-1}(u)$ increases from 0 to 1 as u increases from 0 to ∞. Hence, if $F_{n-1}\left(\frac{n}{n-1}\,u\right)$ is practically unity, i.e. for $\frac{n}{n-1}\,u$ numerically large, the upper percentage points of u_n may be approximated by the normal integral

$$(9)\qquad \begin{aligned} \int_{u_n}^{\infty} dF(u_n) &= \frac{n}{\sqrt{2\pi}}\int_{u_n}^{\infty} \exp\left[-\frac{1}{2}\,\frac{n}{n-1}\,u_n^2\right]\frac{\sqrt{n}}{\sqrt{n-1}}\, du_n \\ &= \frac{n}{\sqrt{2\pi}}\int_{\sqrt{n/(n-1)}\,u_n}^{\infty} \exp\left[-\frac{t^2}{2}\right] dt. \end{aligned}$$

Formula (9) was found to be particularly useful in checking the higher probabilities in Table II.

The cumulative distribution functions (8) may be put into another form by setting

$$u_r = \frac{1}{r}\,v_r\,; \qquad r = 2, 3, \cdots, n\ .$$

Then $F_n(u)$ becomes

$$(10)\qquad \begin{aligned} F_n(u) = \frac{\sqrt{n}}{(\sqrt{2\pi})^{n-1}}\int_0^{nu}\int_0^{v_n}\int_0^{v_{n-1}}\cdots\int_0^{v_4}\int_0^{v_3} \\ \cdot \exp\left[-\frac{1}{2}\sum_{i=2}^{n}\frac{v_i^2}{i(i-1)}\right] dv_2\, dv_3 \cdots dv_n. \end{aligned}$$

Define the following functions:

$$H_1(x) = 1,$$

$$H_2(x) = \sqrt{2}\int_0^x \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\cdot\frac{t^2}{2\cdot 1}\right] H_1(t)\,dt,$$

$$\vdots$$

$$H_n(x) = \sqrt{\frac{n}{n-1}}\int_0^x \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\cdot\frac{t^2}{n(n-1)}\right] H_{n-1}(t)\,dt.$$

Hence, the probability that the difference between the extreme and the mean in samples of n from a normal population is less than $u\sigma$ is given by the alternative forms

$$P\{u_n \leq u\sigma\} = F_n(u) = H_n(nu).$$

Of course, $H_n(nu) \to 1$ as $u \to \infty$ for any given n.

In the November 1945 issue of *Biometrika*, Godwin [13] arrived at a series of functions closely related to the $H_r(x)$ in connection with the distribution of the mean deviation in samples of n from a normal parent. In Godwin's work, he defines functions $G_r(x)$ which are related to the $H_r(x)$ by the equation

$$(2\pi)^{r/2} H_{r+1}(x) = G_r(x).$$

The $G_r(x)$ functions were computed by H. O. Hartley [15] for $r = 2, 3, \cdots 9$ only. Computations on the functions $F_n(u)$, i.e. (8), were well under way by the author before Godwin's article on the mean deviation appeared. The $H_r(x)$ or $G_r(x)$ can be used to obtain both the distribution of the difference between the extreme and mean and also the probability integral of the mean deviation. Indeed, it is believed that these functions may have a useful place in tabulating distributions of order statistics.

6. Tabulation of the distribution function, $F_n(u)$.

The tabulation of the $F_n(u)$ with ordinary computing equipment is quite laborious. However, a table model computing machine was used initially to obtain the $F_n(u)$ for $n = 2$ to $n = 15$ using formulae (8) and a numerical quadrature process.

In view of the possible general usefulness of the $H_r(x)$, these functions were also computed as a sample problem on a high-speed computing device, the ENIAC (Electronic Numerical Integrator and Computor) of the Ballistic Re-

[4] The author suggested the problem of tabulating the functions $F_n(u)$ or $H_n(nu)$ to the Computing Laboratory of the Ballistic Research Laboratories in the fall of 1945; however, due to problems of higher priority, these functions were not computed on the ENIAC until March, 1948.

search Laboratories of the Ordnance Department.[4] In this connection, the $H_r(u)$ have been computed for $r = 2$ to $r = 25$ at the Ballistic Research Laboratories. For $n = 2$, the functions $H_r(x)$ were computed to nine decimal places of accuracy on the ENIAC and at $n = 25$ about five decimal places of accuracy were obtained. In Table II we have tabulated $F_n(u)$ or $H_n(nu)$, i.e. the probability integral of the extreme minus the mean, at intervals of $u = .05\sigma$. Values computed on the table model computing machine agreed to five decimal places at $n = 15$ with values from the ENIAC. Percentage Points of the distribution are given in Table III and the moment constants may be found in Table IV. Moment constants for $n = 60$, 100, 200, 500 and 1000 were obtained by use of McKay's formulae [11] (which relate the semi-invariants of $x_n - \bar{x}$ with those of x_n) and Tippetts moments [5] for the largest observation x_n.

TABLE III

Percentage Points for Extreme Minus Mean

n	90%	95%	99%	99.5%
2	1.163	1.386	1.821	1.985
3	1.497	1.738	2.215	2.396
4	1.696	1.941	2.431	2.618
5	1.835	2.080	2.574	2.764
6	1.939	2.184	2.679	2.870
7	2.022	2.267	2.761	2.952
8	2.091	2.334	2.828	3.019
9	2.150	2.392	2.884	3.074
10	2.200	2.441	2.931	3.122
11	2.245	2.484	2.973	3.163
12	2.284	2.523	3.010	3.199
13	2.320	2.557	3.043	3.232
14	2.352	2.589	3.072	3.261
15	2.382	2.617	3.099	3.287
16	2.409	2.644	3.124	3.312
17	2.434	2.668	3.147	3.334
18	2.458	2.691	3.168	3.355
19	2.480	2.712	3.188	3.375
20	2.500	2.732	3.207	3.393
21	2.519	2.750	3.224	3.409
22	2.538	2.768	3.240	3.425
23	2.555	2.784	3.255	3.439
24	2.571	2.800	3.269	3.453
25	2.587	2.815	3.282	3.465

TABLE IV

Moment Constants for Extreme Minus Mean

n	Mean	Std. Dev.	α_3	α_4
2	.5642	.4263	.9953	3.8692
3	.8463	.4755	.8296	3.7135
4	1.0294	.4916	.7675	3.6717
5	1.1630	.4974	.7372	3.6560
6	1.2672	.4993	.7165	3.6511
7	1.3522	.4991	.7042	3.6503
8	1.4236	.4979	.6959	3.6518
9	1.4850	.4962	.6900	3.6546
10	1.5388	.4943	.6857	3.6582
11	1.5864	.4923	.6827	3.6622
12	1.6292	.4902	.6804	3.6663
13	1.6680	.4881	.6788	3.6705
14	1.7034	.4861	.6777	3.6746
15	1.7359	.4841	.6770	3.6787
20	1.867	.475	.677	3.700
60	2.319	.436	.699	3.801
100	2.508	.418	.712	3.855
200	2.746	.395	.737	3.932
500	3.037	.368	.771	4.033
1000	3.241	.350	.794	4.105

7. Relation between the distribution of the largest minus the mean of all n observations and the largest minus the mean of the remaining n-1 items. The following relation is of interest concerning these two statistics:

Let

$$u_n = x_n - \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$= \frac{1}{n}\{(n-1)\,x_n - x_1 - x_2 - \cdots - x_{n-1}\}.$$

Let

$$v_n = x_n - \frac{x_1 + x_2 + \cdots + x_{n-1}}{n-1}$$

$$= \frac{1}{n-1}\{(n-1)\,x_n - x_1 - x_2 - \cdots - x_{n-1}\}.$$

Hence,

$$v_n = \frac{n}{n-1}\,u_n$$

or

$$P(v_n \leq t_0) = P\left(\frac{n}{n-1} u_n \leq t_0\right) = P\left\{u_n \leq \frac{n-1}{n} t_0\right\},$$

i.e. the probability integral of the largest minus the mean of the other observations may be obtained by interpolation on the distribution of the largest minus the mean of all n items in the sample.

8. The distribution of S_n^2/S^2 and S_1^2/S^2. As indicated in the Summary, we proposed the sample criterion

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k, \qquad \bar{x}_n = \frac{1}{n-1} \sum_{i=1}^{n-1} x_i,$$

for testing the significance of the largest observation and the criterion

$$\frac{S_1^2}{S^2} = \frac{\sum_{i=2}^{n} (x_i - \bar{x}_1)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \leq k, \qquad \bar{x}_1 = \frac{1}{n-1} \sum_{i=2}^{n} x_i,$$

for testing whether the smallest observation is outlying. We now find the probability distribution of S_n^2/S^2; hence, also that of S_1^2/S^2.

Returning to the density function

$$dF(\eta_2, \eta_3, \cdots, \eta_n) = \frac{n!}{(\sqrt{2\pi})^{n-1}} \exp\left[-\frac{1}{2} \sum_{i=2}^{n} \eta_i^2\right] d\eta_2\, d\eta_3 \cdots d\eta_n$$

of Section 5, we make the polar transformation

$$\begin{aligned}
\eta_2 &= r \sin \theta_n \sin \theta_{n-1} \cdots \sin \theta_4 \sin \theta_3, \\
\eta_3 &= r \sin \theta_n \sin \theta_{n-1} \cdots \sin \theta_4 \cos \theta_3, \\
\eta_4 &= r \sin \theta_n \sin \theta_{n-1} \cdots \cos \theta_4, \\
&\;\;\vdots \\
\eta_{n-1} &= r \sin \theta_n \cos \theta_{n-1}, \\
\eta_n &= r \cos \theta_n .
\end{aligned} \tag{11}$$

Now

$$\sum_{i=2}^{n} \eta_i^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = r^2$$

and

$$\sum_{i=2}^{n-1} \eta_i^2 = \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2 = r^2 \sin^2 \theta_n .$$

Hence,

$$\sin^2 \theta_n = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}.$$

The Jacobian of the above transformation is

$$r^{n-2} \sin^{n-3}\theta_n \sin^{n-4}\theta_{n-1} \cdots \sin^3 \theta_6 \sin^2 \theta_5 \sin \theta_4,$$

and since $0 \leq r \leq \infty$

$$dF(\theta_n, \theta_{n-1}, \cdots, \theta_5, \theta_4, \theta_3) \tag{12}$$
$$= \frac{n!}{(2\pi)^{(n-1)/2}} 2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right) \sin^{n-3} \theta_n \cdots \sin^2 \theta_5 \sin \theta_4 \, d\theta_n \cdots d\theta_5 \, d\theta_4 \, d\theta_3.$$

Since the restrictions on the η_i are

$$\eta_2 \geq 0, \qquad \sqrt{\frac{r}{r-2}}\,\eta_r \geq \eta_r - 1, \qquad r \geq 3,$$

we have

$$\tan \theta_n \cos \theta_{n-1} = \frac{\eta_{n-1}}{\eta_n}, \qquad n \geq 4,$$

or

$$\tan \theta_n \leq \sqrt{\frac{n}{n-2}} \sec \theta_{n-1}, \qquad n \geq 4,$$

and

$$0 \leq \theta_3 \leq \frac{\pi}{3}.$$

Thus, letting $K_n = \frac{n!}{(2\pi)^{(n-1)/2}} 2^{(n-3)/2} \Gamma\left(\frac{n-1}{2}\right)$, we see that

$$K_n \int_0^{\pi/3} \int_0^{l_3} \cdots \int_0^{l_{n-2}} \int_0^{l_{n-1}} \sin^{n-3} \theta_n \cdots \sin^2 \theta_5 \sin \theta_4 \, d\theta_n \cdots d\theta_4 \, d\theta_3 = 1, \tag{13}$$

where $l_r = \tan^{-1} \sqrt{\frac{r+1}{r-1}} \sec \theta_r$.

Upon reversing the order of integration (the variable limits are monotonic) we get for $n = 3$

$$K_3 \int_0^{\pi/3} d\theta_3 = 1,$$

so that

$$P(\theta_3 \leq \theta) = K_3 \int_0^{\theta} d\theta_3 \qquad 0 \leq \theta \leq M_3 = \tan^{-1} \sqrt{3 \cdot 1}. \tag{14}$$

When $n = 4$, we obtain

$$K_4 \int_0^{m_4} \int_0^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4 + K_4 \int_{m_4}^{M_4} \int_{L_4}^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4 = 1$$

where

$$m_r = \tan^{-1} \sqrt{\frac{r}{r-2}}, M_r = \tan^{-1} \sqrt{r(r-2)} \text{ and } L_r = \sec^{-1} \sqrt{\frac{r-2}{r}} \tan \theta_r,$$

so that

$$\text{(15a)} \quad P(\theta_4 \leq \theta) = \frac{K_4}{K_3} \int_0^{\theta} \sin \theta_4 \, d\theta_4 \qquad \text{when } 0 \leq \theta \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}$$

and

$$\text{(15b)} \quad P(\theta_4 \leq \theta) = \frac{K_4}{K_3} \int_0^{m_4} \sin \theta_4 \, d\theta_4 + K_4 \int_{m_4}^{\theta} \int_{L_4}^{\pi/3} \sin \theta_4 \, d\theta_3 \, d\theta_4$$

$$\text{when } m_4 = \tan^{-1} \sqrt{\frac{4}{2}} \leq \theta \leq M_4 = \tan^{-1} \sqrt{4 \cdot 2}.$$

When $n = 5$, we get,

$$K_5 \int_0^{m_5} \int_0^{m_4} \int_0^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5 + K_5 \int_0^{m_5} \int_{m_4}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5$$

$$+ K_5 \int_{m_5}^{M_5} \int_{L_5}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5 = 1$$

$\left(\text{where } L_4 = \sec^{-1} \sqrt{\frac{2}{4}} \tan \theta_4 \text{ is to be taken as 0 whenever } \theta_4 \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}\right)$ so that

$$\text{(16a)} \quad P(\theta_5 \leq \theta) = \frac{K_5}{K_4} \int_0^{\theta} \sin^2 \theta_5 \, d\theta_5 \qquad \text{when } 0 \leq \theta \leq m_5 = \tan^{-1} \sqrt{\frac{5}{3}}$$

and

$$\text{(16b)} \quad P(\theta_5 \leq \theta) = \frac{K_5}{K_4} \int_0^{m_5} \sin^2 \theta_5 \, d\theta_5 + K_5 \int_{m_5}^{\theta} \int_{L_5}^{M_4} \int_{L_4}^{\pi/3} \sin^2 \theta_5 \sin \theta_4 \, d\theta_3 \, d\theta_4 \, d\theta_5$$

where $m_5 = \tan^{-1} \sqrt{\frac{5}{3}} \leq \theta \leq M_5 = \tan^{-1} \sqrt{5 \cdot 3}$,
and we put $L_4 = \sec^{-1} \sqrt{\frac{2}{4}} \tan \theta_4 = 0$ whenever $\theta_4 \leq m_4 = \tan^{-1} \sqrt{\frac{4}{2}}$.

For a sample of n items

(17a)
$$P(\theta_n \leq \theta) = \frac{n}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_0^\theta \sin^{n-3} \theta_n \, d\theta_n$$
$$= \frac{n}{2} I_{\sin^2\theta}\left(\frac{n-2}{2}, \frac{1}{2}\right) \qquad \text{when } 0 \leq \theta \leq \tan^{-1} \sqrt{\frac{n}{n-2}}$$[5]

and

(17b)
$$P(\theta_n \leq \theta) = \frac{n}{2} I_{n/(2(n-1))}\left(\frac{n-2}{2}, \frac{1}{2}\right)$$
$$+ K_n \int_{m_n}^{\theta} \int_{L_n}^{M_{n-1}} \int_{L_{n-1}}^{M_{n-2}} \cdots \int_{L_4}^{\pi/2} \sin^{n-3} \theta_n \cdots \sin \theta_4 \, d\theta_3 \, d\theta_4 \cdots d\theta_n$$

for

$$m_n = \tan^{-1} \sqrt{\frac{n}{n-2}} \leq \theta \leq M_n = \tan^{-1}\sqrt{n(n-2)}$$

where $I_x(p, q)$ is K. Pearson's Incomplete Beta Function Ratio [19]. It is to be understood in (17) that

$$L_i = \sec^{-1} \sqrt{\frac{i-2}{i}} \tan \theta_i \quad \text{for} \quad i = 4, 5, \cdots, n-1$$

is to be taken as zero when $\theta_i \leq \tan^{-1} \sqrt{\frac{i}{i-2}}$.

Percentage points for the sample statistic

$$\sin^2 \theta_n = \frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1} (x_1 - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_1 - \bar{x})}$$

or the statistic S_1^2/S_2 are given in Table I and were obtained by inverse interpolation on the tabulation of the probability integral (17) above. Percentage points for the Pearson and Chandra Sekar statistics, $T_n = \frac{x_n - \bar{x}}{s}$ or $T_1 = \frac{\bar{x} - x_1}{s}$ (where $s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$), are given in Table IA. The statistics S_n^2/S^2 and T_n are related by the formula

$$\frac{S_n^2}{S^2} = 1 - \frac{T_n^2}{n-1}.$$

[5] It has been noted that (17a) gives a good approximation to (17b) when $\theta \geq \tan \sqrt{\frac{n}{n-2}}$ provided we are interested in the important practical region $P \leq .10$, at least for $n \leq 25$.

The statistic T_n (or T_1) is easier to compute than S_n^2/S^2 (or S_1^2/S^2). The tabulation of the multiple integral (17) was carried out on the Bell Relay Computors at the Ballistic Research Laboratories.

9. The distribution of $S^2_{n-1,n}/S^2$ and $S^2_{1,2}/S^2$. As indicated in the Summary, the proposed criterion for judging the significance of the two largest observations is

$$S^2_{n-1,n}/S^2 = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \le k \qquad \text{where } \bar{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i ,$$

and that for testing the two smallest observations is

$$S^2_{1,2}/S^2 = \frac{\sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \le k \qquad \text{where } \bar{x}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} x_i .$$

From the preceding section, we note that

$$\sum_{i=2}^{n} \eta_i^2 = r^2, \qquad \sum_{i=2}^{n-2} \eta_i^2 = r^2 \sin^2\theta_n \sin^2\theta_{n-1}.$$

Hence,

$$\sin^2\theta_n \sin^2\theta_{n-1} = \frac{\sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{18}$$

so that if we find the distribution of

$$\sin^2\theta_n \sin^2\theta_{n-1} = \sin^2\Delta_n , \text{ say,}$$

then we have the distribution of $S^2_{n-1,n}/S^2$ and hence also that of $S^2_{1,2}/S^2$, i.e.

$$P\{\sin^2\Delta_n \le k\} = P\{\Delta_n \le \sin^{-1}\sqrt{k}\}. \tag{19}$$

Returning to the multiple integral (13), let

$$\sin\Delta_n = \sin\theta_n \sin\theta_{n-1},$$

$$\Delta_i = \theta_i, \qquad 3 \le i \le n-1.$$

The Jacobian of this transformation is given by

$$\frac{\partial(\theta_n, \cdots, \theta_3)}{\partial(\Delta_n, \cdots, \Delta_3)} = \frac{\cos\Delta_n}{\sqrt{\sin^2\Delta_{n-1} - \sin^2\Delta_n}}.$$

The limits of integration for Δ_n are given by

$$0 \le \Delta_n \le \sin^{-1}\frac{\sqrt{n}\sin\Delta_{n-1}}{\sqrt{2(n-1) - (n-2)\sin^2\Delta_{n-1}}}$$

and, of course, those for $\Delta_{n-1}, \cdots, \Delta_3$ are the same as the limits for $\theta_{n-1}, \cdots, \theta_3$ respectively. Hence, substituting in (13), we obtain

$$
K_n \int_0^{\pi/3} \int_0^{\tan^{-1}\sqrt{4/2}\sec\Delta_3} \cdots \int_0^{\tan^{-1}\sqrt{n-1/n-3}\sec\Delta_{n-2}} \cdot \int_0^{\sin^{-1}\frac{\sqrt{n}\sin\Delta_{n-1}}{\sqrt{2(n-1)-(n-2)\sin^2\Delta_{n-1}}}} \frac{\sin^{n-3}\Delta_n \sin^{n-4}\Delta_{n-1} \cdots \sin^2\Delta_5 \sin\Delta_4 \cos\Delta_n \, d\Delta_n \cdots d\Delta_3}{\sin^{n-3}\Delta_{n-1}\sqrt{\sin^2\Delta_{n-1} - \sin^2\Delta_n}} = 1. \tag{20}
$$

Reversing the order of integration, we have

$$
K_n \int_0^{\sin^{-1}\sqrt{\frac{n(n-3)}{(n-1)(n-2)}}} \int_{\sin^{-1}\frac{\sqrt{2(n-1)}\sin\Delta_n}{\sqrt{n+(n-2)\sin^2\Delta_n}}}^{\tan^{-1}\sqrt{(n-1)(n-3)}} \int_{\sec^{-1}\sqrt{\frac{n-3}{n-1}}\tan\Delta_{n-1}}^{\tan^{-1}\sqrt{(n-2)(n-4)}} \cdots \int_{\sec^{-1}\sqrt{2/4}\tan\Delta_4}^{\pi/3} \frac{\sin^{n-3}\Delta_n \sin^{n-4}\Delta_{n-1} \cdots \sin\Delta_4 \cos\Delta_n \, d\Delta_3 \cdots d\Delta_n}{\sin^{n-3}\Delta_{n-1}\sqrt{\sin^2\Delta_{n-1} - \sin^2\Delta_n}} = 1 \tag{21}
$$

$\left(\text{for } \Delta_i \leq \tan^{-1}\sqrt{\frac{i}{i-2}}\right.$, then $\sec^{-1}\sqrt{\frac{i-2}{i}}\tan\Delta_i$ is to be put equal to zero where $\left. i \geq 4\right)$ so that for $n = 4$,

$$
P(\Delta_4 \leq \Delta) = K_4 \int_0^{\Delta} \int_{\sin^{-1}\frac{\sqrt{3}\sin\Delta_3}{\sqrt{2+\sin^2\Delta_3}}}^{\pi/3} \frac{\sin\Delta_4 \cos\Delta_4 \, d\Delta_3 \, d\Delta_4}{\sin\Delta_3 \sqrt{\sin^2\Delta_3 - \sin^2\Delta_4}} \tag{22}
$$

where $0 \leq \Delta \leq \sin^{-1}\sqrt{\frac{2}{3}}$,

and for $n = 5$,

$$
P(\Delta_5 \leq \Delta) = K_5 \int_0^{\Delta} \int_{\sin^{-1}\frac{\sqrt{4\cdot 2}\sin\Delta_5}{\sqrt{5+3\sin^2\Delta_5}}}^{\tan^{-1}\sqrt{4\cdot 2}} \cdot \int_{\sin^{-1}\sqrt{2/4}\tan\Delta_4}^{\pi/3} \frac{\sin^2\Delta_5 \cos\Delta_5 \, d\Delta_3 \, d\Delta_4 \, d\Delta_5}{\sin\Delta_4 \sqrt{\sin^2\Delta_4 - \sin^2\Delta_5}} \tag{23}
$$

where $0 \leq \Delta \leq \sin^{-1}\sqrt{\frac{5}{6}}$, etc.

We remark that an obvious extension of the above principles should lead to the distributions of

$$
S^2_{n-2,n-1,n}/S^2 \quad \text{and} \quad S^2_{1,2,3}/S^2,
$$
$$
S^2_{n-3,n-2,n-1,n}/S^2 \quad \text{and} \quad S^2_{1,2,3,4}/S^2,
$$

etc. although the tabulation of such probability integrals may be exceedingly difficult.

The problem of tabulating the probability integral (21) involves a double quadrature process and has been carried out on the Bell Relay Computors at the Ballistic Research Laboratories for $n = 4$ to $n = 20$, inclusive. Table V gives some useful percentage points for these sample sizes.

TABLE V

Table of Percentage Points for $\frac{S^2_{n-1,n}}{S^2}$ *or* $\frac{S^2_{1,2}}{S^2}$

n	1%	2.5%	5%	10%
4	.0000	.0002	.0008	.0031
5	.0035	.0090	.0183	.0376
6	.0186	.0349	.0565	.0921
7	.0440	.0708	.1020	.1479
8	.0750	.1101	.1478	.1994
9	.1082	.1492	.1909	.2454
10	.1415	.1865	.2305	.2863
11	.1736	.2212	.2666	.3226
12	.2044	.2536	.2996	.3552
13	.2333	.2836	.3295	.3843
14	.2605	.3112	.3568	.4106
15	.2859	.3367	.3818	.4345
16	.3098	.3603	.4048	.4562
17	.3321	.3822	.4259	.4761
18	.3530	.4025	.4455	.4944
19	.3725	.4214	.4636	.5113
20	.3909	.4391	.4804	.5269

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S^2_{n-1,n} = \sum_{i=1}^{n-2} (x_i - \bar{x}_{n-1,n})^2 \quad \text{where} \quad \bar{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i$$

$$S^2_{1,2} = \sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2 \quad \text{where} \quad \bar{x}_{1,2} = \frac{1}{n-2}\sum_{i=3}^{n} x_i$$

10. Comment on the distribution of $S^2_{1,n}/S^2$. In connection with the distribution of the statistic

$$\frac{S^2_{1,n}}{S^2} = \frac{\sum_{i=2}^{n-1} (x_i - \bar{x}_{1,2})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \quad \text{where} \quad \bar{x}_{1,n} = \frac{1}{n-2}\sum_{i=2}^{n-1} x_i,$$

for testing simultaneously whether the smallest and largest observations are outlying, an investigation indicates that since

$$\sum x_i^2 = n\bar{x}^{-2} + \frac{n}{n-1}(x_n - \bar{x})^2 + \frac{n-1}{n-2}(x_1 - \bar{x}_n)^2 + \frac{n-2}{n-3}(x_{n-1} - \bar{x}_{1,n})^2$$
$$+ \cdots + \frac{3}{2}\left(x_4 - \frac{x_2 + x_3 + x_4}{3}\right)^2 + 2\left(x_3 - \frac{x_2 + x_3}{2}\right)^2$$

then the transformation

$$
\begin{aligned}
\sqrt{2\cdot 1}v_2 &= -x_2 + x_3, \\
\sqrt{3\cdot 2}v_3 &= -x_2 - x_3 + 2x_4, \\
\sqrt{4\cdot 3}v_4 &= -x_2 - x_3 - x_4 + 3x_5, \\
&\;\;\vdots \\
\sqrt{(n-2)(n-3)}v_{n-2} &= -x_2 - x_3 - \cdots - x_{n-2} + (n-3)x_{n-1}, \\
\sqrt{(n-1)(n-2)}v_{n-1} &= -(n-2)x_1 + x_2 + x_3 + \cdots + x_{n-1}, \\
\sqrt{n(n-1)}v_n &= -x_1 - x_2 - x_3 - \cdots - x_{n-1} + (n-1)x_n, \\
\sqrt{n}v_{n+1} &= x_1 + x_2 + \cdots + x_n,
\end{aligned}
\tag{24}
$$

followed by transformations of the type (11) and that of Section 9 may lead to the distribution of $S_{1,n}^2/S^2$. However, the limits of integration do not turn out to be functions of single variables and the task of computing the resulting multiple integral may be rather difficult.

11. Examples on testing outlying observations for rejection. We now turn to the problem of applying our theory to particular practical examples of data which appear to have outlying observations. Apparently, in the following examples there were not sufficient practical or experimental grounds to reject the suspected outliers and hence some statistical judgement became necessary either to support retaining the "outliers" in the sample or leave little doubt that certain of the observations should be questioned.

EXAMPLE 1. Our first example has almost become a classical one as Irwin [3], Rider [2], and other writers on the subject including Chauvenet, Peirce, Gould, etc. (see Rider's survey [2]) all refer to it, applying their various tests. The example consists of a sample of 15 observations of the vertical semi-diameters of Venus made by Lieut. Herndon in 1846 and is given in William Chauvenet's, *A Manual of Spherical and Practical Astronomy*, II (5th ed., 1876), p. 562. The individual residuals or deviations from the mean are:

−0.30″	0.48	0.63	−0.22	0.18
−0.44	−0.24	−0.13	−0.05	0.39
1.01	0.06	−1.40	0.20	0.10

Arranging the observations in increasing order of magnitude, we have:

−1.40″	−0.24	−0.05	0.18	0.48
−0.44	−0.22	0.06	0.20	0.63
−0.30	−0.13	0.10	0.39	1.01

and it is seen that two of the residuals, -1.40 and 1.01, appear to be outliers. Rider [2] indicates that the above observations have been referred to by previous writers as "residuals"; nevertheless their sum is 0.27, so that the sample mean, $\bar{x} = .018$. Let us apply the exact test, i.e. T_1 of Pearson and Chandra Sekar or S_1^2/S^2 as developed in Section 8 for a single outlier to the least observation, -1.40. We find $x_1 = -1.40$, $\bar{x} = .018$ and $s = .532$ (alternatively, we find $S^2 = 4.2496$ using all 15 observations and $S_1^2 = 2.0953$ which is based on 14 observations, the suspected outlier -1.40 not being included). Further, $T_1 = \dfrac{\bar{x} - x_1}{s} = \dfrac{.018 + 1.40}{.532} = 2.665$ (or $S_1^2/S^2 = 0.4931$) and from Table IA (or Table I) we see that $0.01 \leq P \leq 0.025$ so that we would reject the observation -1.40 when using the 5% level of significance. Having rejected -1.40, we now have left a sample of 14 observations and test the greatest one, i.e. 1.01. For T_n based on the remaining 14 observations, we have $n = 14$, $x_n = 1.01$, $\bar{x} = .119$ and $s = .387$ (alternatively, for the new sums of squares, we find $S_n^2 = 1.2409$ leaving out 1.01 and $S^2 = 2.0953$ including the observation 1.01). Hence, $T_n = \dfrac{x_n - \bar{x}}{s} = \dfrac{1.01 - .119}{.387} = 2.302$ (or $S_n^2/S^2 = 0.5922$) and from Table IA (or I), we find P slightly less than .10, so that we decide to retain the observation 1.01.

It would have been interesting nevertheless to see whether or not the test $S_{1,n}^2/S^2$ would have rejected simultaneously the observations -1.40 and 1.01 if percentage points for the distribution of this statistic were available.

It is of interest to remark that for this particular example Irwin [3, page 245], using the difference between the first two individuals divided by an estimate of σ, i.e. $\dfrac{x_2 - x_1}{\sigma}$, concluded also that -1.40 but not 1.01 should be rejected. In testing *both* of these observations, Irwin used the single biased estimate for σ,

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = = .5326 \qquad (\text{assuming } \bar{x} = 0),$$

based on all 15 observations. It is a mere coincidence, of course, that for this example Irwin's test gives the same result as the exact test T_1 or the test based on the ratio S_1^2/S^2. In this connection, Irwin rightly calls attention to the fact that in dealing with a sample of only 15 observations the standard deviation of the sample is a very unreliable estimate of the population standard deviation.

It is remarked that here we would, of course, hesitate to apply the test $\dfrac{\bar{x} - x_1}{\sigma}$ to the observation -1.40 as we do not have available and accurate estimate of σ from past data.

Example 2. The following ranges (horizontal distances from gun muzzle to point of impact) were obtained in firing projectiles from a weapon at a constant angle of elevation and at the same weight of charge of propellant powder:

Distances in yards

4782	4420
4838	4803
4765	4730
4549	4833

It is desired to know whether the projectiles exhibit uniformity in ballistic behavior or if some of the ranges, such as 4549 and 4420, are not consistent with the others.

Arranging the distances or ranges in increasing order of magnitude,

4420	4782
4549	4803
4730	4833
4765	4838

we suspect the presence of two outliers, i.e. 4420 and 4549. Having no available knowledge of σ from past data for this example, an intuitively efficient test to apply would be that of Section 9, i.e. $S_{1,2}^2/S^2$.
We find

$$\frac{S_{1,2}^2}{S^2} = \frac{\sum_{i=3}^{8} (x_i - \bar{x}_{1,2})^2}{\sum_{=1}^{8} (x_i - \bar{x})^2} = .054$$

which is significant at the .01 level (Table V) and consequently we would judge the distances 4420 and 4549 yds. as being unusually low.

As a matter of interest and as a recommended temporary practical expedient for testing several "outliers", consider for example the last seven of the above ordered observations,

4549	4803
4730	4833
4765	4838
4782	

and apply the exact test, S_1^2/S^2, to the smallest observation, 4549. We find $S_1^2/S^2 = .145$ so that $.01 < P < .025$ from Table I and we should thus reject 4549 from the sample of seven. Moreover, we should now surely reject 4420 as being outlying, arriving at the same result we had for the test $S_{1,2}^2/S^2$. Thus, as a general temporary expedient in testing for "outliers" one could rank the observations, and apply the tests S_1^2/S^2 (or S_n^2/S^2) and $S_{1,2}^2/S^2$ (or $S_{1,n}^2/S^2$), thus working from the "inside" observations of the ranked sample in order to establish consistency of the observations.

12. Additional comments. Although we have used a significance level of .05 in the examples, it may be preferable from a practical viewpoint to reject outlying observations only at a lower level, such as .01 or .005.

Extensions of the ideas for testing outlying observations presented in this paper may lead to efficient sample criteria for testing the significance of various numbers of high, low, or simultaneously high and low sample values. However, the mathematical details would probably be complicated. In this connection, it is remarked nevertheless that the advent of high-speed computing devices may have considerable bearing on establishing experimentally any probability distribution. That is to say high-speed electronic computing devices could probably be programmed to generate random numbers with frequencies equal to those of the normal (or any other) distribution, to compute various functions (such as ratios in this paper) of sample values, etc., and establish frequency distributions to a desired order of accuracy.

13. Acknowledgement. The author is greatly indebted to Prof. C. C. Craig under whose most competent guidance this work was carried out. Indeed, the stimulus and encouragement received by the author as a result of Prof. Craig's interest in the problem were paramount in orienting various phases of the subject matter and in accomplishing the results given herein. In connection with the computing, debts of gratitude are owed to several members of the staff of the Ballistic Research Laboratories. It is desired to express appreciation to Col. Leslie E. Simon, Director of the Ballistic Research Laboratories, for recognizing the importance and desirability of carrying out the computations on high-speed computing devices. In this connection, appreciation is also expressed to Dr. L. S. Dederick, Chief, Computing Laboratory, Ballistic Research Laboratories. The programming of the $H_n(u)$ functions on the Electronic Numerical Integrator and Computor was done by Dr. Derrick Lehmer of the University of California, who was with the Computing Laboratory, BRL, during the latter part of World War II and Miss Ruth Lichterman. The author is particularly indebted to Dr. Franz Alt, Dr. Bernard Dimsdale, Miss R. Lichterman, Mr. John Holberton, Miss H. Marks, Mr. F. Spence, and others of the Computing Laboratory during the period the functions $H_n(\chi)$ were computed on the ENIAC. The computing of the distribution of the statistics S_n^2/S^2 and $S_{n-1,n}^2/S^2$ was done on the Bell Relay Computors under the direction of Mr. J. O. Harrison, Mrs. M. Masincup and Mr. E. Cushen. The author also desires to express appreciation to Miss Helen J. Coon for considerable computation and checking carried out on a table-model computing machine.

REFERENCES

[1] E. S. PEARSON AND C. CHANDRA SEKAR, "The efficiency of statistical tools and a criterion for the rejection of outlying observations", *Biometrika*, Vol. 28 (1936), pp. 308–320.

[2] P. R. RIDER, "Criteria for rejection of observations", *Washington University Studies*—New Series, Science and Technology—No. 8, St. Louis (1933).

[3] J. O. Irwin, "On a criterion for the rejection of outlying observations", *Biometrika*, Vol. 17 (1925), pp. 238–250.

[4] "Student", *Biometrika*, Vol. 19 (1927), pp. 151–164.

[5] L. H. C. Tippett, "The extreme individuals and the range of samples taken from a normal population", *Biometrika*, Vol. 17 (1925), pp. 151–164.

[6] E. S. Pearson, "A further note on the distribution of range in samples taken from a normal population", *Biometrika*, Vol. 18 (1926), pp. 173–194.

[7] *Tables for Statisticians and Biometricians, Part II*, edited by Karl Pearson, pp. CX–CXIX.

[8] E. S. Pearson, "The percentage limits for the distribution of range in samples from a normal population", *Biometrika*, Vol. 24 (1932), pp. 404–417.

[9] H. O. Hartley, "The range in random samples", *Biometrika*, Vol. 32 (1942), pp. 334–348.

[10] E. S. Pearson and H. O. Hartley, "The probability integral of the range in samples of n observations from a normal population", *Biometrika*, Vol. 32 (1942), pp. 301–310.

[11] A. T. McKay, "The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe", *Biometrika*, Vol. 27 (1935), pp. 466–471.

[12] W. R. Thompson, "On a criterion for the rejection of observations and the distribution of the ratio of the deviation to the sample standard deviation", *Annals of Math. Stat.*, Vol. 6 (1935), pp. 214–219.

[13] H. J. Godwin, "On the distribution of the estimate of mean deviation obtained from samples from a normal population", *Biometrika*, Vol. 33 (1945), pp. 254–256.

[14] W.P.A., *Tables of Probability Functions*, Vols. I and II, New York, N.Y. (1942).

[15] H. O. Hartley, "Note on the calculation of the distribution of the estimate of mean deviation in normal samples", *Biometrika*, Vol. 33 (1945), pp. 257–258.

[16] "Tables of the probability integral of the mean deviation in normal samples", *Biometrika*, Vol. 33 (1945), pp. 259–265.

[17] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Phil. Trans. Roy. Soc.* (London), Vol. 231 (1933), pp. 289–337.

[18] E. S. Pearson and H. O. Hartley, "Tables of the probability integral of the studentized range,", *Biometrika*, Vol. 33 (1943), pp. 89–99.

[19] K. Pearson, *Tables of the Incomplete Beta-Function*, published by the Biometrika Office, University College, London.

[20] K. R. Nair, "The distribution of the extreme deviate from the sample mean and its studentized form," *Biometrika*, Vol. 35 (1948), pp. 118–144.

DISTRIBUTION OF THE CIRCULAR SERIAL CORRELATION COEFFICIENT FOR RESIDUALS FROM A FITTED FOURIER SERIES[1,2]

By R. L. Anderson and T. W. Anderson[3]

North Carolina State College and Columbia University

CONTENTS

Summary. In this paper the observations are considered to be normally distributed with constant variance and means consisting of linear combinations of certain trigonometric functions. The likelihood ratio criterion for testing the independence of the observations against the alternatives of circular serial correlation of a given lag is found to be a function of the circular serial correlation coefficient for residuals from the fitted Fourier series (Section 4). The exact distribution (Section 5), the moments (Section 6), and approximate distributions

[1] Included in Cowles Commission Papers, New Series, No. 42.

[2] Presented to the meeting of the Institute of Mathematical Statistics at New York, December 30, 1947.

[3] Fellow of the John Simon Guggenheim Memorial Foundation; Research Consultant of the Cowles Commission for Research in Economics.

(Section 7) are given for the cases of greatest interest. From these results significance levels have been found (Section 3). The use of these levels is indicated (Section 2), and an example of their use is given (Section 3).

1. Introduction. Two mathematical models have been used extensively in time-series analysis. In one model the observation is the sum of a "systematic part" and a random error. The cyclical properties of this model result from the cyclical properties of the systematic part, which is usually taken to be a short Fourier series. The stochastic element is superimposed on the non-stochastic part, and the error at one time point does not affect a later observation. The other model is the stochastic difference equation or "autoregressive model." An observation is the sum of a linear function of previous observations and a random element. The cyclical properties follow from the properties of the difference equation (i.e., the linear combination of observations), but are disturbed by the random disturbance that is integrated into the system. A more general model can be constructed that includes both of the two mentioned. The observation can be taken as a linear combination of past observations and Fourier terms plus a random element.

In this paper, the linear combination will be only a multiple of some preceding observation. For lag 1, the model is of the form

$$x_i - \mu_i = \rho(x_{i-1} - \mu_{i-1}) + u_i, \qquad i = 1, 2 \cdots, N, \tag{1}$$

where $x_0 \equiv x_n$ and $\mu_0 \equiv \mu_n$. In (1), the $\{x_i\}$ are the N observations; the $\{u_i\}$ are N random disturbances, each assumed normally and independently distributed with zero mean and variance σ^2; the means $\{\mu_i\}$ are linear combinations of some of the N functions of i: $\cos \frac{2\pi ig}{N}$ and $\sin \frac{2\pi ih}{N}$. For N odd, $g = 0, 1, \cdots, \frac{1}{2}(N-1)$; $h = 1, \cdots, \frac{1}{2}(N-1)$. For N even, $g = 0, 1, \cdots, \frac{1}{2}N$; $h = 1, \cdots, \frac{1}{2}N - 1$. Hence,

$$\mu_i = \sum_{g'} \alpha_{g'} \cos \frac{2\pi ig'}{N} + \sum_{h'} \beta_{h'} \sin \frac{2\pi ih'}{N}, \tag{2}$$

where g' and h' run over certain values of the ranges of g and h, respectively. Let K' be the number of terms in (2). Usually the constant term, α_0, is included (in this case $g = 0$ and $\cos \frac{2\pi ig}{N} = 1$). Of the N trigonometric functions available, the terms in (2) are usually chosen so that terms with certain periods are included and terms with other periods are excluded. It should be noted that (1) defines a circular model.

The sample estimates of $\alpha_{g'}$ and $\beta_{h'}$ are the usual regressions of x_i on $\cos \frac{2\pi ig'}{N}$ and $\sin \frac{2\pi ih'}{N}$, respectively. Because of the orthogonality of these trigonometric terms, the estimates are

(3)
$$
\begin{aligned}
a_{g'} &= \sum_{i=1}^{N} x_i \cos \frac{2\pi i g'}{N} \Big/ \frac{N}{2}, \qquad g \neq 0, \tfrac{1}{2}N,\\
b_{h'} &= \sum_{i=1}^{N} x_i \sin \frac{2\pi i h'}{N} \Big/ \frac{N}{2},\\
a_0 &= \sum_{i=1}^{N} x_i \Big/ N,\\
a_{\frac{1}{2}N} &= \sum_{i=1}^{N} x_i \cos \pi i \Big/ N = \sum_{i=1}^{N} (-1)^i x_i \Big/ N.
\end{aligned}
$$

The fitted series is

(4)
$$m_i = \sum_{g'} a_{g'} \cos \frac{2\pi i g'}{N} + \sum_{h'} b_{h'} \sin \frac{2\pi i g'}{N}.$$

where the sums on g' and h' are over the ranges in (2).

The serial correlation coefficient suitable for this model is

(5)
$$R = \frac{\sum_{i=1}^{N} (x_i - m_i)(x_{i-1} - m_{i-1})}{\sum_{i=1}^{N} (x_i - m_i)^2},$$

where $m_0 \equiv m_n$. This statistic can be used to estimate ρ, or it can be used to test hypotheses about ρ. In fact, for the circular model this statistic leads to the best tests [3].

It is hoped that the mathematical model studied in this paper can be used in the treatment of certain problems in economic time series. For example, the seasonal variation in a series of data may be considered as a "systematic part" made up of trigonometric components. In the next section we discuss in a more detailed way how the use of this model may arise in the field of economics.

We have considered circular serial correlation, although in most statistical problems it is non-circular serial correlation that is involved. The reason for treating the circular case is the inherent mathematical simplicity. The circular coefficient and Fourier series of the type (2) are "naturally" related. The relevant fact is that the vectors

$$\left(\cos \frac{2\pi g}{N}, \cos \frac{4\pi g}{N}, \cdots, \cos \frac{2N\pi g}{N}\right) \text{ and } \left(\sin \frac{2\pi h}{N}, \sin \frac{4\pi h}{N}, \cdots, \sin \frac{2N\pi h}{N}\right)$$

are characteristic vectors of the matrix of the quadratic form in $(x_i - m_i)$ of the numerator of R. For this reason the distribution and significance points are easily obtained.

In the usual applications the circular coefficient can be used even if the hypothesis alternative to independence of observations is non-circular serial correla-

tion. The circular coefficient may not have as good power against non-circular alternatives as non-circular coefficients, such as

$$(6) \qquad \frac{\sum_{i=2}^{N} (x_i - m_i)(x_{i-1} - m_{i-1})}{\sum_{i=1}^{N} (x_i - m_i)^2}.$$

However, the difference between these two statistics is $(x_1 - m_1)(x_N - m_N)/\Sigma(x_i - m_i)^2$, and it can be shown that this converges stochastically to zero (as N increases and ρ remains fixed).

2. The use of fitted Fourier series. A linear combination of trigonometric terms may be used as a regression function when there is a "systematic part" (or "trend") that is periodic. For instance, it may be reasonable to assume that a series of agricultural data has a systematic component with certain periodicities due to variation in weather. Then one may ask whether this regression function "explains" all of the interrelations in the series.

An example taken from agricultural economics is a development of that given by Koopmans [8]. Suppose p_t and q_t are the price and supply, respectively, of a given farm product at time t. Let $Q_t^{(d)}$ be the quantity demanded at time t if $p_t = P$, and $Q_t^{(s)}$ be the quantity supplied at time t if $p_{t-L} = P$, where P is an arbitrarily selected point of reference on the price scale, serving to define the Q's. Let the market equations be defined as follows:

$$(7) \qquad p_t - P = -\beta(q_t - Q_t^{(d)}) + u_t,$$

$$(8) \qquad q_t - Q_t^{(s)} = \delta(p_{t-L} - P) + v_t,$$

where u and v are random disturbances. The first equation expresses the price depressing tendency of an abnormally large supply; the second expresses the supply-stimulating influence of abnormally high prices L time units earlier (the time between planning the product and selling it). We can substitute from (7) at time $(t - L)$ into (8) and obtain

$$(9) \qquad q_t - Q_t^{(s)} = \rho(q_{t-L} - Q_{t-L}^{(d)}) + w_t,$$

which is of the form (1) for general lag L ($i - 1$ is replaced by $t - L$) if $Q_t^{(s)} - \rho Q_{t-L}^{(d)} = \mu_t - \rho\mu_{t-L}$. Now we may wish to test the null hypothesis, $H_0 : \rho = 0$. If we assume that our alternative hypothesis is $H_a : \rho > 0$, we can test the null hypothesis by use of the positive tail of the distribution of R. Similarly for $H_a : \rho < 0$, we would use the negative tail of the distribution of R. In other cases, if we believe $\rho \neq 0$, we might wish to estimate ρ.

It is of particular interest to consider using the Fourier series for seasonal variation. The most important cases are given below with indications of the appropriate tables of significance points for testing the hypothesis $\rho = 0$. (a) *Annual data.* Here only a constant is fitted; this is the sample mean. The tables

given in [2] or [5] are to be used. (b) *Semi-annual data.* To "correct" for variation of period two we fit a constant and $\cos \pi t = (-1)^t$. The table given in Section 3 for $P = 2$ is to be used. (c) *Quarterly data.* The four terms to be fitted are 1, $\cos \pi t = (-1)^t$, $\cos \frac{\pi t}{2}$, and $\sin \frac{\pi t}{2}$. The table given in Section 3 for $P = 2$ and 4 is to be used. (d) *Bimonthly data.* The six terms to be fitted are 1, $\cos \pi t$, $\cos \frac{2\pi t}{3}$, $\sin \frac{2\pi t}{3}$, $\cos \frac{\pi t}{3}$, and $\sin \frac{\pi t}{3}$. The table given in Section 3 for $P = 2, 3$, and 6 is to be used. (e) *Monthly data.* The twelve terms to be fitted are 1, $\cos \frac{\pi t}{6}$, $\sin \frac{\pi t}{6}$, $\cos \frac{\pi t}{3}$, $\sin \frac{\pi t}{3}$, $\cos \frac{\pi t}{2}$, $\sin \frac{\pi t}{2}$, $\cos \frac{2\pi t}{3}$, $\sin \frac{2\pi t}{3}$, $\cos \frac{5\pi t}{6}$, $\sin \frac{5\pi t}{6}$, and $\cos \pi t = (-1)^t$. The table given in Section 3 for $P = 2, 12/5, 3, 4, 6$, and 12 is to be used. It is assumed here that the data are given for each time interval in a certain number of years. Then the residuals are the same as the residuals taken from means for each month or season. That is, if the data are monthly, one may compute the sample means for January, February, etc., and residuals are to be taken from the corresponding monthly means. The fitted Fourier coefficients are certain linear functions of these means.

3. Tables of significance points of R.

3.1. *Significance points of R using a seasonal trend for annual, semi-annual, bimonthly, and monthly data.* The calculations of significance points of R (lag 1 only) have been subdivided according to the number of terms included in the estimating equations, m_i. The significance points for only a constant in m_i have been tabulated in [2] and [5]. Since the main use for m_i equations involving sine and cosine terms seems to be for semi-annual, quarterly, bimonthly, and monthly data, for which N is even, the results presented in this paper are for N even. Then we will have all of the sine and cosine terms in pairs except for $\cos \pi i = (-1)^i$ and the constant term. We shall find it convenient to refer to the period $P_{g'} = N/g'$ or $P_{h'} = N/h'$ of the terms in (2).

We have calculated significance points R' exact to 3 decimal places, for $Pr\{R > R'\} = \alpha = .01, .05, .95$, and $.99$. The values of R' corresponding to $\alpha = .01$ and $.05$ are usually indicated as the positive significance points and those corresponding to $\alpha = .95$ and $.99$, the negative significance points. In all of these cases, except for annual data, the distribution of R is symmetrical. Hence only the positive significance points need be given, since the negative points are simply the corresponding positive points with opposite sign; that is, $R'(.95) = -R'(.05)$, $R'(.99) = -R'(.01)$.

The significance points were calculated from the exact distribution of R given in Section 5 for all N up to the values where the approximate significance points using an Incomplete Beta distribution (Section 7) were the same as the exact significance points. The Incomplete Beta significance points were used

up to the value of N for which a normal approximation was satisfactory. For some of the results, the normal points became sufficiently accurate to be used following the exact points.

The values of R' are given in Table 1 except for (a), for the following values of N:

(a) *Annual data*—see the tables in [2] or [5].

(b) *Semi-annual data* ($P = 2$): $N = 6(2)60$. The exact points were needed for N through 10($\alpha = .05$) and N through 22 ($\alpha = .01$). The normal points could be used for $N = 60$ ($\alpha = .05$) but were still too large by .003 for $N = 60$ ($\alpha = .01$)

(c) *Quarterly data* ($P = 2, 4$): $N = 8(4)100$. The exact points were needed for N through 20 ($\alpha = .05$) and N through 32 ($\alpha = .01$). The normal points were adequate for all N above 20 ($\alpha = .05$) but were still too large by .001 for $N = 100$ ($\alpha = .01$).

(d) *Bimonthly data* ($P = 2, 3, 6$): $N = 12(6)150$. The exact points were needed for N through 24 ($\alpha = .05$) and N through 30 ($\alpha = .01$). Again the normal points were adequate for all N above 24 ($\alpha = .05$) but were still too large by .0005 for $N = 150$ ($\alpha = .01$).

(e) *Monthly data* ($P = 2, 12/5, 3, 4, 6, 12$): $N = 24(12)300$. The exact points were needed for $N = 24$ ($\alpha = .05$) and $N = 24, 36$ ($\alpha = .01$). The normal points were adequate for $N > 24$ ($\alpha = .05$) and $N > 300$ ($\alpha = .01$).[4]

Significance points for the Incomplete Beta approximation (See Section 7) are tabulated in terms of $2p$ and $2q$. The values of $2p$ and $2q$ are the same when $\mu_1'(R) = 0$; for (c), (d), and (e) above these values are simply $N - 3$, $N - 5$, and $N - 11$, respectively. Hence, for two-tailed significance points for these cases, the ordinary correlation tables can be used with $N - 3$, $N - 5$, and $N - 11$ degrees of freedom, respectively. Also, our one-tailed significance points can be approximated by use of the 10% and 2% significance points for the ordinary correlation coefficient. 10%, 5%, 2%, 1%, and 0.1% two-tailed significance points have been tabulated by Fisher and Yates [6]. These significance points are accurate to three decimal places for the serial correlation coefficients as follows:[5]

(c) $n = N - 3$ degrees of freedom: $N \geq 24$ ($\alpha = .05$); $N \geq 36$ ($\alpha = .01$),

(d) $n = N - 5$ degrees of freedom: $N \geq 24$ ($\alpha = .05$); $N \geq 30$ ($\alpha = .01$),

(e) $n = N - 11$ degrees of freedom: $N \geq 24$ ($\alpha = .05$ and $\alpha = .01$), where α is the one-tailed significance point. For semi-annual data (b), $2p = 2q = \dfrac{N^2 - 3N + 4}{N - 4}$, which is not an integer for $N > 12$. When $N = 12$, $2p = 2q = 14$, for which the ordinary correlation significance point is adequate for $\alpha = .05$.

---

[4] It should be noted that for (c), (d), and (e), an approximation given by Cochran [4] is easily computed and is more accurate than the normal approximation for the $\alpha = .01$ significance points.

[5] In [6] n is 2 less than the number of pairs used in computing the ordinary correlation coefficient when the sample means are first subtracted.

Details of computing techniques using the exact distribution are given by R. L. Anderson [1] for computing values of R' when $\mu_i = 0$.

3.2. *Significance points of R for other single-period trends.* Significance points have also been obtained for $P = 3$, $P = 4$, $P = 6$, and $P = 12$, for which $K' = 3$.

TABLE 1

*Exact significance points, R', for different fitted series**

	$P = 2$			$P = 2,4$			$P = 2, 3, 6$			$P = 2, 12/5, 3, 4, 6, 12$	
$N \diagdown \alpha$	.05	.01	$N \diagdown \alpha$	.05	.01	$N \diagdown \alpha$	.05	.01	$N \diagdown \alpha$	.05	.01
6	.495	.499	8	.636	.693	12	.592	.744	24	.441	.592
8	.484	.607	12	.515	.661	18	.442	.592	36	.323	.445
10	.453	.601	16	.439	.582	24	.369	.504	48	.267	.371
12	.426	.572	20	.388	.523	30	.323	.445	60	.233	.325
14	.402	.544	24	.351	.478	36	.291	.403	72	.209	.293
16	.382	.519	28	.323	.441	42	.267	.371	84	.191	.268
18	.364	.496	32	.300	.414	48	.248	.346	96	.177	.249
20	.348	.476	36	.282	.391	54	.233	.325	108	.166	.234
22	.334	.458	40	.267	.371	60	.220	.308	120	.157	.221
24	.321	.442	44	.254	.354	66	.209	.293	132	.149	.210
26	.310	.427	48	.243	.338	72	.200	.280	144	.142	.200
28	.300	.414	52	.233	.325	78	.191	.268	156	.136	.192
30	.290	.402	56	.224	.313	84	.184	.258	168	.131	.184
32	.282	.390	60	.216	.302	90	.177	.249	180	.126	.178
34	.274	.380	64	.209	.293	96	.172	.241	192	.122	.172
36	.266	.370	68	.202	.284	102	.166	.234	204	.118	.166
38	.260	.361	72	.197	.276	108	.161	.227	216	.115	.162
40	.254	.353	76	.191	.268	114	.157	.221	228	.111	.157
42	.248	.345	80	.186	.261	120	.153	.215	240	.108	.153
44	.242	.338	84	.182	.255	126	.149	.210	252	.105	.149
46	.237	.331	88	.177	.249	132	.145	.205	264	.103	.146
48	.233	.324	92	.173	.243	138	.142	.200	276	.101	.142
50	.228	.318	96	.170	.238	144	.139	.196	288	.099	.140
52	.224	.313	100	.166	.234	150	.136	.192	300	.097	.136
54	.220	.307									
56	.216	.302									
58	.212	.297									
60	.209	.292									

* P = Periods Used in Fitted Series.

In these cases, the distribution of R is asymmetrical. The Incomplete Beta approximation is symmetrical for $P = 3$, with $2p = 2q = N - 2$, even though the exact distribution is not.

The significance points for these single-period trends are given in Table 2.

The exact distribution was required to compute the $\alpha = .01$ and .99 significance points for N through 48 in all cases and also for most cases with $\alpha = .05$ and .95. For $N > 48$, the Cochran approximation [4] gave the same results as the Incomplete Beta approximation. Since this Cochran approximation can be computed more rapidly, it should be used if other significance points are desired. The normal approximation is not recommended because it is less accurate than the Cochran approximation and requires almost as much calculation. For $\alpha = .01$ and .99, the significance points using the normal approximation were too large (in absolute value) by from .0005 to .001 for the last entries in Table 2. The two-

TABLE 2

Exact significance points, R', for single periods > 2

$P = 3$					$P = 6$				
N	α				N	α			
	.99	.95	.05	.01		.99	.95	.05	.01
6	−.970	−.854	.496	.500					
12	−.690	−.522	.475	.619	12	−.766	−.651	.296	.506
18	−.558	−.409	.392	.526	18	−.630	−.509	.277	.440
24	−.480	−.348	.340	.463	24	−.540	−.427	.254	.393
30	−.428	−.309	.304	.417	30	−.482	−.373	.236	.359
36	−.389	−.280	.277	.382	36	−.438	−.335	.220	.332
42	−.360	−.257	.256	.356	42	−.403	−.306	.207	.311
48	−.336	−.240	.240	.334	48	−.375	−.283	.197	.294
54	−.316	−.226	.226	.316	54	−.352	−.264	.188	.279
60	−.300	−.214	.214	.300	60	−.333	−.248	.180	.266
66	−.286	−.204	.204	.286	66	−.316	−.235	.173	.255
72	−.274	−.195	.195	.274	72	−.301	−.224	.167	.246
78	−.263	−.187	.187	.263	78	−.288	−.214	.161	.237
84	−.254	−.181	.181	.254	84	−.277	−.205	.156	.229
90	−.245	−.175	.175	.245	90	−.267	−.197	.151	.222
96	−.237	−.169	.169	.237	96	−.258	−.190	.147	.216
102	−.230	−.164	.164	.230	102	−.250	−.184	.143	.210
108	−.224	−.159	.159	.224	108	−.242	−.178	.140	.205
114	−.218	−.155	.155	.218	114	−.235	−.173	.137	.200
120	−.212	−.151	.151	.212	120	−.229	−.168	.134	.195
126	−.207	−.147	.147	.207	126	−.223	−.163	.131	.191
132	−.202	−.144	.144	.202	132	−.218	−.159	.128	.187
138	−.198	−.141	.141	.198	138	−.213	−.155	.125	.183
144	−.194	−.138	.138	.194	144	−.208	−.152	.123	.180
150	−.190	−.135	.135	.190	150	−.203	−.148	.121	.177

TABLE 2—*Continued*

$P = 4$					$P = 12$				
N	α				N	α			
	.99	.95	.05	.01		.99	.95	.05	.01
8	−.889	−.768	.503	.637	12	−.778	−.671	.096	.245
12	−.742	−.608	.420	.585	24	−.555	−.444	.197	.330
16	−.643	−.502	.369	.522	36	−.447	−.348	.188	.298
20	−.576	−.441	.333	.474	48	−.383	−.293	.175	.270
24	−.519	−.396	.306	.437	60	−.339	−.257	.163	.249
28	−.477	−.361	.285	.407	72	−.307	−.231	.153	.231
32	−.445	−.334	.268	.383	84	−.283	−.212	.145	.217
36	−.418	−.312	.253	.363	96	−.263	−.196	.138	.206
40	−.395	−.293	.241	.345	108	−.247	−.183	.132	.196
44	−.375	−.277	.230	.330	120	−.233	−.173	.126	.187
48	−.358	−.264	.221	.317	132	−.221	−.164	.121	.180
52	−.343	−.252	.213	.305	144	−.211	−.156	.117	.173
56	−.330	−.242	.206	.294	156	−.202	−.149	.113	.167
60	−.319	−.233	.199	.285	168	−.194	−.143	.110	.162
64	−.308	−.225	.193	.277	180	−.187	−.138	.107	.157
68	−.298	−.218	.188	.269	192	−.181	−.133	.104	.153
72	−.289	−.211	.183	.262	204	−.175	−.128	.101	.149
76	−.281	−.205	.178	.255	216	−.170	−.124	.099	.145
80	−.274	−.199	.174	.249	228	−.165	−.121	.097	.141
84	−.267	−.194	.170	.243	240	−.161	−.117	.094	.138
88	−.261	−.189	.166	.238	252	−.157	−.114	.092	.135
92	−.255	−.184	.162	.233	264	−.153	−.111	.091	.132
96	−.249	−.180	.159	.228	276	−.149	−.109	.089	.130
100	−.244	−.176	.156	.223	288	−.146	−.106	.087	.127
108	−.234	−.169	.150	.215	300	−.143	−.104	.086	.125
120	−.221	−.160	.143	.205					
132	−.210	−.152	.136	.196					
144	−.201	−.145	.131	.187					

tailed significance points cannot be obtained from the ordinary correlation tables except for $P = 3$.

3.3. *Example of use of significance points.* As an example of the use of these significance points, R', we shall consider the following data [17] on the receipts of butter (in units of 1,000,000 pounds) at five markets (Boston, Chicago, San Francisco, Milwaukee, and St. Louis). The figures in parentheses are deviations from the average of the given months over the 3 years.

Month	Year			Total T_i	Average
	1935	1936	1937		
Jan.	48.9(2.4)	48.3(1.8)	42.4(−4.1)	139.6	46.5
Feb.	43.4(−0.6)	47.1(3.1)	41.4(−2.6)	131.9	44.0
March	43.8(−4.6)	52.4(4.0)	49.0(0.6)	145.2	48.4
April	50.8(−1.5)	55.3(3.0)	50.8(−1.5)	156.9	52.3
May	67.6(1.6)	64.7(−1.3)	65.8(−0.2)	198.1	66.0
June	83.7(0.7)	79.5(−3.5)	85.9(2.9)	249.1	83.0
July	82.7(10.7)	62.6(−9.4)	70.6(−1.4)	215.9	72.0
Aug.	60.8(4.8)	51.3(−4.7)	55.8(−0.2)	167.9	56.0
Sept.	55.4(3.6)	51.0(−0.8)	49.1(−2.7)	155.5	51.8
Oct.	48.4(−1.0)	54.0(4.6)	45.7(−3.7)	148.1	49.4
Nov.	37.7(−4.5)	45.2(3.0)	43.8(1.6)	126.7	42.2
Dec.	41.0(−3.2)	44.9(0.7)	46.7(2.5)	132.6	44.2
Total	664.2(8.4)	656.3(0.5)	647.0(−8.8)	1967.5	655.8
Average	55.35(0.70)	54.69(0.04)	53.92 (−0.73)	163.96	54.65

We assume that the trend is composed of the 12 terms having periods that divide 12. We shall test the null hypothesis that the deviations from the trend are independently distributed against the alternative that there is positive serial correlation. The fitted series is of the form

$$m_i = b_0^* + \sum_{j=1}^{5}\left(b_{2j-1}^* \cos\frac{\pi i j}{6} + b_{2j}^* \sin\frac{\pi i j}{6}\right) + b_{11}^* \cos \pi i; \tag{10}$$

here we find it convenient to use the notation, $b_0^*, b_1^*, \cdots, b_{11}^*$, for the coefficients (with a different relationship between the subscripts and the trigonometric functions than in (4)). We find that the m_i are simply the average receipts given for each month in the above table (46.5, 44.0, $\cdots$, 44.2). Hence the deviations $(x_i - m_i)$ are given by the figures in parentheses (2.4, −0.6, $\cdots$, 2.5). The calculated lag 1 circular serial correlation coefficient is

$$\begin{aligned} R_0 &= \frac{(2.4)(-0.6) + (-0.6)(-4.6) + \cdots + (1.6)(2.5) + (2.5)(2.4)}{(2.4)^2 + (-0.6)^2 + \cdots + (2.5)^2} \\ &= \frac{232.18}{474.51} = 0.489. \end{aligned} \tag{11}$$

Entering Table 1 for P = 2, 12/5, 3, 4, 6, and 12 and N = 36, we find that R' (.05) = 0.323 and R' (.01) = .445. Hence, at either the 5% or 1% level the null hypothesis of zero serial correlation ($\rho = 0$) is to be rejected (against the alternative single-tail hypothesis, $\rho > 0$). If we had been interested in the two-

tailed alternative hypothesis, $\rho \neq 0$, we would use the ordinary correlation tables with $N - 11 = 25$ degrees of freedom and we would find that for the two-tailed test $R'(.01) = 0.487$. Our value is significant at the 5% level and barely significant at the 1% level.

The values of b^* in (10) are computed as follows

$$
\begin{aligned}
b_0^* &= \sum_{i=1}^{12} T_i/36, \\
b_{2j-1}^* &= \sum_{i=1}^{12} T_i \cos \frac{\pi i j}{6} \Big/ 18, \\
b_{2j}^* &= \sum_{i=1}^{12} T_i \sin \frac{\pi i j}{6} \Big/ 18, \\
b_{11}^* &= \sum_{i=1}^{12} T_i \cos \pi i/36.
\end{aligned} \tag{12}
$$

The computed values of b_0^* to b_{11}^* are 54.65, -14.82, -2.02, 6.60, 1.23, -3.98, 0.30, 2.21, 1.73, -0.61, 0.60, 0.15, respectively. However, it is not necessary to compute these values in order to obtain m_i. The problem of estimating the variances of these b's will be discussed in Section 4.

4. Testing the hypothesis of lack of serial correlation.

4.1. *Statement of the problem.* Consider the N random variables $u_1, \cdots, u_N$, each normally and independently distributed with mean 0 and variance σ^2. Define the N variables $x_1, \cdots, x_N$ by the equations

$$x_i - \mu_i = \rho(x_{i-L} - \mu_{i-L}) + u_i \qquad (i = 1, \cdots, N), \tag{13}$$

where

$$x_{-j} \equiv x_{N-j}, \; \mu_{-j} \equiv \mu_{N-j} \qquad (j = 0, 1, \cdots, N-1) \tag{14}$$

and μ_i is the linear combination of trigonometric functions given in (2). If L and N are relatively prime (in particular, if $L = 1$), the Jacobian of the transformation from $\{u_i\}$ to $\{x_i\}$ is $1 - \rho^N$, and the probability density of $\{x_i\}$ is

$$\frac{1 - \rho^N}{(2\pi\sigma^2)^{\frac{1}{2}N}} e^{-\frac{1}{2}Q/\sigma^2}, \tag{15}$$

where $Q = (1 + \rho^2) \sum_{i=1}^{N} (x_i - \mu_i)^2 - 2\rho \sum_{i=1}^{N} (x_i - \mu_i)(x_{i-L} - \mu_{i-L})$. If $L = 1$, the covariance between x_i and x_j is $\sigma^2[\rho^{|i-j|} + \rho^{N-|i-j|}]/[(1 - \rho^N)(1 - \rho^2)]$. If $L = q\alpha$ and $N = p\alpha$, where p, q, and α are positive integers and q and p are relatively prime, then the Jacobian is $(1 - \rho^p)^\alpha$ and the density of $\{x_i\}$ is

$$\frac{(1 - \rho^p)^\alpha}{(2\pi\sigma^2)^{\frac{1}{2}N}} e^{-\frac{1}{2}Q/\sigma^2}. \tag{16}$$

We shall now obtain the likelihood ratio test of the hypothesis $H_0 : \rho = 0$ on the basis of a sample consisting of one observation on each x_i.

4.2. *Preliminary transformations.* We shall find it convenient to express μ_i in terms of fixed variates ϕ_{ij}, having certain properties. Later we will verify that the ϕ's are simply constant multiples of the trigonometric terms in (2). We suppose now that

$$\mu_i = \sum_{j=1}^{K'} \phi_{ij}\gamma_j \qquad (i = 1, \cdots, N), \tag{17}$$

where $K' < N$, the $\{\gamma_j\}$ are parameters, and the ϕ_{ij} are known functions of i and j satisfying

$$\phi_{i-L,j} + \phi_{i+L,j} = 2\lambda_{Lj}\phi_{ij} \qquad (i = 1, \cdots, N; \quad j = 1, \cdots, K'), \tag{18}$$

$$\sum_{i=1}^{N} \phi_{ij}\phi_{ik} = \delta_{jk} \qquad (j, k = 1, \cdots, K'), \tag{19}$$

$$\phi_{-i,j} = \phi_{N-i,j} \qquad (i = 0, 1, \cdots, N - 1), \tag{20}$$

and δ_{jk} is the Kronecker delta. Let

$$m_i = \sum_{j=1}^{K'} \phi_{ij} c_j, \tag{21}$$

where

$$c_j = \sum_{i=1}^{N} x_i \phi_{ij}. \tag{22}$$

Then by usual regression theory we have

$$\sum_{i=1}^{N} (x_i - m_i)\phi_{ij} = 0, \tag{23}$$

$$\sum_{i=1}^{N} (x_i - \mu_i)^2 = \sum_{i=1}^{N} (x_i - m_i)^2 + \sum_{j=1}^{K'} (c_j - \gamma_j)^2 \tag{24}$$

because c_j is the least squares estimate of γ_j. Let us evaluate

$$\begin{aligned} {}_L\bar{C} &= \sum_{i=1}^{N} (x_i - \mu_i)(x_{i-L} - \mu_{i-L}) \\ &= \sum_{i=1}^{N} [(x_i - m_i) + (m_i - \mu_i)][(x_{i-L} - m_{i-L}) + (m_{i-L} - \mu_{i-L})] \\ &= \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) + \sum_{i=1}^{N} \sum_{j=1}^{K'} \phi_{i-L,j}(c_j - \gamma_j)(x_i - m_i) \\ &\quad + \sum_{i=1}^{N} \sum_{j=1}^{K'} \phi_{ij}(c_j - \gamma_j)(x_{i-L} - m_{i-L}) \\ &\quad + \sum_{i=1}^{N} \sum_{j,k=1}^{K} \phi_{ik}\phi_{i-L,j}(c_k - \gamma_k)(c_j - \gamma_j). \end{aligned} \tag{25}$$

Call the first term on the right hand side of (25) ${}_LC$. In view of (20) the next two terms are

$$\sum_{j=1}^{K'} \sum_{i=1}^{N} (x_i - m_i)(\phi_{i-L,j} + \phi_{i+L,j})(c_j - \gamma_j). \tag{26}$$

This is seen to be zero by consideration of (18) and (23). The last term can be written

$$\frac{1}{2} \sum_{i=1}^{N} \sum_{j,k=1}^{K'} (\phi_{ik}\phi_{i-L,j} + \phi_{i+L,j}\phi_{ik})(c_k - \gamma_k)(c_j - \gamma_j) = \sum_{j=1}^{K'} \lambda_{Lj}(c_j - \gamma_j)^2 \tag{27}$$

by use of (18), (19), and (20). Thus

$${}_L\bar{C} = \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) + \sum_{j=1}^{K'} \lambda_{Lj}(c_j - \gamma_j)^2. \tag{28}$$

It follows that

$$\begin{aligned} Q = (1 + \rho^2) \sum_{i=1}^{N} (x_i - m_i)^2 - 2\rho \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) \\ + \sum_{j=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{Lj})(c_j - \gamma_j)^2. \end{aligned} \tag{29}$$

We can complete the matrix $\Phi = (\phi_{ij})$ so that Φ is an N-th order square matrix with elements satisfying (18), (19), and (20). If we make the transformation

$$x_i = \sum_{j=1}^{N} \phi_{ij} c_j \qquad (i = 1, \cdots, N), \tag{30}$$

then

$$\sum_{i=1}^{N} (x_i - m_i)^2 = \sum_{j=K'+1}^{N} c_j^2, \tag{31}$$

$$\sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}) = \sum_{j=K'+1}^{N} \lambda_{Lj} c_j^2. \tag{32}$$

4.3. *The likelihood ratio criterion.* To obtain the likelihood ratio test of the hypothesis $H_0 : \rho = 0$ against alternative hypotheses $H_a : \rho \neq 0$, we divide the maximum of the likelihood assuming H_0 by the maximum of the likelihood assuming H_a. It is clear from (15) and (29) that if H_0 is true, the maximum likelihood estimates of γ_j and σ^2 are c_j and

$$s_0^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m_i)^2, \tag{33}$$

respectively. If H_a is true, the maximum likelihood estimate of γ_j is c_j. To state the maximum likelihood estimates of σ^2 and ρ under H_a it is convenient to define ${}_LR$, the sample serial coefficient of lag L, as

$${}_LR = \frac{1}{N s_0^2} \sum_{i=1}^{N} (x_i - m_i)(x_{i-L} - m_{i-L}). \tag{34}$$

Then the maximum likelihood estimate of σ^2 under H_a is

$$s^2 = s_0^2(1 + \hat{\rho}^2 - 2\hat{\rho}\,{}_LR), \tag{35}$$

where $\hat{\rho}$ is the maximum likelihood estimate of ρ and satisfies

$${}_LR(1 + \hat{\rho}^N) - \hat{\rho}(1 + \hat{\rho}^{N-2}) = 0, \tag{36}$$

if L and N are relatively prime and satisfies

$${}_LR(1 + \hat{\rho}^p) - \hat{\rho}(1 + \hat{\rho}^{p-2}) = 0, \tag{37}$$

if $L = q\alpha$, $N = p\alpha$, and p and q are relatively prime.

Upon substituting these estimates into the likelihood function we find that the likelihood ratio criterion is

$$\lambda = \frac{(1 + \hat{\rho}^2 - 2\hat{\rho}\,{}_LR)^{\frac{1}{2}N}}{1 - \hat{\rho}^N}, \tag{38}$$

if L and N are relatively prime and

$$\lambda = \left[\frac{(1 + \hat{\rho}^2 - 2\hat{\rho}\,{}_LR)^{\frac{1}{2}p}}{1 - \hat{\rho}^p}\right]^\alpha, \tag{39}$$

if $L = q\alpha$, $N = p\alpha$ and p and q are relatively prime. The maximum likelihood estimate of ρ is the root of (36) or (37) that makes (38) or (39), respectively, a minimum. It should be noticed that throughout this section ρ could be replaced by $1/\rho$ (and changing σ^2 by a factor $1 + \rho^2$). To make the maximum likelihood estimate unique, we require that $|\hat{\rho}| \leq 1$. It can be shown that there exists one and only one root of (36) or (37) that satisfies this requirement and minimizes λ. (There is a peculiarity to this solution in that if N is odd, $L = 1$, and ${}_1R < -1 + 2/N$, then $\hat{\rho} = -1$ is the root minimizing λ). In any case, λ is a function of ${}_LR$. We have shown that for $0 < {}_LR < 1$, it is a monotonic decreasing function; and for $-1 < {}_LR < 0$, it is a monotonic increasing function. A critical region defined by $\lambda \leq \lambda_0$ can, therefore, be defined by ${}_LR \leq R_1 < 0$ and $0 < R_2 \leq {}_LR$. (The probability that ${}_LR = -1$ or $+1$ is 0.) Thus we can use ${}_LR$ to test the null hypothesis $H_0 : \rho = 0$ instead of the likelihood ratio criterion (against one-sided alternatives they are equivalent). The strongest justification for the use of ${}_LR$ in testing $H_0 : \rho = 0$ is that for circular distributions the uniformly most powerful tests against one-sided alternatives and the B_1 test against two-sided alternatives are given in terms of inequalities on ${}_LR$ [3].

We can also use ${}_LR$ as an estimate of ρ. In fact, ${}_LR$ is asymptotically a root of (36) or (37). This is proved by showing that ${}_LR(1 + {}_LR^N) - {}_LR(1 + {}_LR^{N-2}) = {}_LR^{N-1}(1 - {}_LR^2)$ converges stochastically to zero. We shall use ${}_LR$ both to estimate ρ and to test hypotheses about this parameter.[6]

Now we shall define ϕ_{ij} used in Section 4.2 in terms of the trigonometric terms indicated in Section 1. In the rest of the paper we shall let the index g run from

---

[6] W. J. Dixon [5] arrived at ${}_LR$ as the maximum likelihood estimate for μ_i a constant by neglecting the Jacobian in (15).

0 to $\frac{1}{2}N$ for N even and from 0 to $\frac{1}{2}(N - 1)$ for N odd; we let the index h run from 1 to $\frac{1}{2}N - 1$ for N even and from 1 to $\frac{1}{2}(N - 1)$ for N odd. We shall use a prime to denote an index running over those values corresponding to fitted terms and a double prime to denote an index running over those values corresponding to terms not fitted.

Let the N trigonometric functions of i, namely $\cos \dfrac{2\pi i g}{N}$ and $\sin \dfrac{2\pi i h}{N}$, be numbered from 1 to N such that the fitted terms are numbered from 1 to K' and the non-fitted terms from $K' + 1$ to N. According to this numbering we define ϕ_{i1} to ϕ_{iN} as

(40) $$\phi_{ij} = \sqrt{\frac{2}{N}} \cos \frac{2\pi i g}{N},$$

or

(41) $$\phi_{ij} = \sqrt{\frac{2}{N}} \sin \frac{2\pi i h}{N}.$$

Defined this way, the ϕ_{ij} satisfy (18) and (19) and (20). It can be shown by using the addition formulas for sines and cosines that

(42) $$\lambda_{L,j} = \cos \frac{2\pi L f}{N},$$

where $f = g$ or $f = h$ depending on whether j refers to a term (40) or (41). We shall assume that the numbering of trigonometric functions is such that

(43) $$\lambda_{L,K'+1} \geq \lambda_{L,K'+2} \geq \cdots \geq \lambda_{L,N}.$$

It can easily be seen that (2) is of the form (17) except that $\alpha_{g'}$ and $\beta_{h'}$ must be multiplied by $\sqrt{\frac{1}{2}N}$ unless $g' = 0$ or $\frac{1}{2}N$ and by $\sqrt{N}$ for $g' = 0, \frac{1}{2}N$ to obtain γ_j. The regression coefficients $a_{g'}$ and $b_{h'}$ are similarly related to the c_j.

It can be seen from (29) that the a_f and b_f are independently distributed with variance $\frac{1}{2}N\sigma^2 \Big/ \left(1 + \rho^2 - 2\rho \cos \dfrac{2\pi L f}{N}\right)$ for $f \neq 0, \frac{1}{2}N$ and variance $N\sigma^2/(1 - \rho)^2$ for $f = 0$ and for $f = \frac{1}{2}N$ if L is even and $N\sigma^2/(1 + \rho)^2$ for $f = \frac{1}{2}N$ if L is odd. In these variance formulas we can estimate σ^2 from (35) using ${}_LR$ for $\hat{\rho}$ and ρ.

5. The exact distribution of ${}_LR$.

5.1. *Introduction.* Under the null hypothesis $H_0 : \rho = 0$ the observations $\{x_i\}$ are normally and independently distributed with variance σ^2 and means $Ex_i = \mu_i$. The variables c_j defined by (22) and (29) are normally and independently distributed with variance σ^2 and means γ_j. For $j > K'$, $\gamma_j = 0$. It follows from (31), (32), (33), and (34) that

(44) $${}_LR = \frac{\sum_{j=K'+1}^{N} \lambda_{Lj} c_j^2}{\sum_{j=K'+1}^{N} c_j^2},$$

where the $\lambda_{L,j}$ are given by (42) corresponding to the $K'' = (N - K')$ trigonometric terms not fitted. Thus to obtain the distribution of ${}_LR$ we need only consider the joint distribution of $\{c_j\}$, $j = K' + 1, \cdots, N$. If H_a is true, the joint density of all the c_j is (15), where

$$Q = (1 + \rho^2)V - 2\rho\,{}_LC + \sum_{j=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{Lj})(c_j - \gamma_j)^2, \tag{45}$$

and

$$V = \sum_{j=K'+1}^{N} c_j^2 \quad \text{and} \quad {}_LC = \sum_{j=K'+1}^{N} \lambda_{Lj} c_j^2 .$$

5.2. *Some special distributions of* ${}_1R = R$. If the constant term $(g' = 0)$ is fitted and the other terms are fitted in pairs $\left(\cos \frac{2\pi i f}{N} \text{ and } \sin \frac{2\pi i f}{N}\right)$, then K' is odd. If N is odd, then K'' is even; the λ_{1j} occur in pairs and we can define λ_k'' as

$$\begin{aligned} \lambda_{1,K'+1} = \lambda_{1,K'+2} = \lambda_1'' > \lambda_{1,K'+3} = \lambda_{1,K'+4} \\ = \lambda_2'' > \cdots > \lambda_{1,N-1} = \lambda_{1N} = \lambda_{\frac{1}{2}K''}'' . \end{aligned} \tag{46}$$

This also holds if N is even and if, in addition to the constant term and paired cosines and sines, we fit $\cos \pi i = (-1)^i (g' = N/2)$. If N is even and we do not fit $\cos \pi i$, we have K'' odd. Then

$$\begin{aligned} \lambda_{1,K'+1} = \lambda_{1,K'+2} = \lambda_1'' > \lambda_{1,K'+3} = \lambda_{1,K'+4} = \lambda_2'' > \cdots > \lambda_{1,N-2} \\ = \lambda_{1,N-1} = \lambda_{\frac{1}{2}(K''-1)}'' > \lambda_{1N} = \lambda_{\frac{1}{2}(K''+1)}'' = -1. \end{aligned} \tag{47}$$

The general expression for the distribution of R in these cases has been found by one of the authors [2]. In this case the cumulative distribution function is 1 minus

$$\begin{gathered} Pr\{R > R'\} = \sum_{k=1}^{m} (-1)^{k+1} \mid V_k \mid (\lambda_k'' - R')^{\frac{1}{2}K''-1}, \\ \lambda_{m+1}'' \le R' \le \lambda_m'', \end{gathered} \tag{48}$$

where V_k is found from a result of Lehmann [9] to be

$$V_k = \frac{2^{\frac{1}{2}(N+1)}}{N} \sin \frac{2\pi f''}{N} \sin \frac{\pi f''}{N} \prod_{j'} \sqrt{(\lambda_k'' - \lambda_{1j'})}, \tag{49}$$

where f'' is such that $\lambda_k'' = \cos \dfrac{2\pi f''}{N}$ and the product on j' is over the K' terms $\lambda_{1j'}$, excluding $\lambda_{1j'} = 1$. Hence, $\lambda_{1j'}$ takes on $K' - 1$ values in $\frac{1}{2}(K' - 1)$ pairs if K' is odd and in $\frac{1}{2}(K' - 2)$ pairs plus a single $\lambda_{1,j'} = -1$ if K' is even. We can also write V_k as

$$V_k = \frac{2^{\frac{1}{2}(N+K')}}{N} \sin \frac{2\pi f''}{N} \sin \frac{\pi f''}{N} \prod_{g' \neq 0} \sqrt{\sin \frac{\pi(g' + f'')}{N} \sin \frac{\pi(g' - f'')}{N}} \cdot \prod_{h'} \sqrt{\sin \frac{\pi(h' + f'')}{N} \sin \frac{\pi(h' - f'')}{N}}. \tag{50}$$

5.3. *Some special distributions of $_LR$ for $L > 1$.* We have noted in (44) above that $\lambda_{L,j} = \cos \frac{2\pi L f''}{N}$, where f'' corresponds to a term not used in the estimation equations for m_i, which was a function of $\left\{\cos \frac{2\pi i g'}{N}, \sin \frac{2\pi i h'}{N}\right\}$. If L, the lag, is relatively prime to N, the distribution is the same as that given above for $L = 1$, except for the re-evaluating of the λ_k''. In the article by R. L. Anderson [2], where only the constant term in m_i was used, the λ_k'' for lag L were exactly the same as the λ_k'' for lag 1. However, this will not be the case for other terms used in m_i. For example, consider lag 2 and N odd with m_i consisting of the constant term plus terms in $\cos \frac{2\pi i}{N}$ and $\sin \frac{2\pi i}{N}$. In this case the λ_k'' for lag 1 are $\left\{\cos \frac{4\pi}{N}, \cos \frac{6\pi}{N}, \cdots, \cos \frac{(N-1)\pi}{N}\right\}$ and the λ_k'' for lag 2 are

$$\left\{\cos \frac{2\pi}{N}, \cos \frac{6\pi}{N}, \cos \frac{8\pi}{N}, \cdots, \cos \frac{(N-1)\pi}{N}\right\}.$$

Next suppose the highest common factor of L and N is α (as before, $L = q\alpha$ and $N = p\alpha$, with p and q relatively prime). In this case

$$\lambda_{L,j} = \cos \frac{2\pi q f''}{p}. \tag{51}$$

Since p and q are relatively prime, the results are the same as for q replaced by 1 and L replaced by α. Each root is repeated α times.

$N = 2L (p = 2)$

If we let $N = 2L$, $\lambda_k'' = \cos \pi k = +1$ or -1. $\lambda'' = +1$ corresponds to these fitted terms in m_i: $\left\{1, \cos \frac{2\pi i g'}{N}, \sin \frac{2\pi i h'}{N}\right\}$ for g', h' even. $\lambda'' = -1$ corresponds to these terms: $\left\{\cos \frac{2\pi i g'}{N}, \sin \frac{2\pi i h'}{N}\right\}$ for g', h' odd. Let $L - n_1$ be the number of terms pertaining to $\lambda'' = +1$ and $L - n_2$ be the number of terms for $\lambda'' = -1$. Then, as in [2], we have the density

$$D(_LR_2) = \frac{(1 - {}_LR_2)^{\frac{1}{2}(n_2-2)}(1 + {}_LR_2)^{\frac{1}{2}(n_1-2)}}{2^{\frac{1}{2}(n_1+n_2)-1}\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)}, \tag{52}$$

where $_LR_2$ was the notation used for lag L and $p = 2$. The cumulative function is the Incomplete Beta function, found by setting $x = \frac{1}{2}(1 - R')$.

$N = 3L (p = 3)$

If we let $N = 3L$, $\lambda_k'' = \cos \frac{2\pi f''}{N} = +1, -\frac{1}{2}$. The fitted terms in m_i corresponding to $\lambda'' = 1$ are $\left\{1, \cos \frac{2\pi i g'}{N}, \sin \frac{2\pi i h'}{N}\right\}$ for $g', h' = 3m$. Similarly, those corresponding to $\lambda'' = -\frac{1}{2}$ have $g', h' = 3m - 1$ or $3m - 2$. Let the number of fitted terms with $\lambda'' = +1$ be $L - n_1$ and with $\lambda'' = -\frac{1}{2}$ be $2L - n_2$. Then

$$D({}_LR_3) = \frac{(1 - {}_LR_3)^{\frac{1}{2}(n_2-2)}(\frac{1}{2} + {}_LR_3)^{\frac{1}{2}(n_1-2)}}{(3/2)^{\frac{1}{2}(n_1+n_2)-1}\beta(\frac{1}{2}n_1, \frac{1}{2}n_2)}, \tag{53}$$

where ${}_LR_3 \geq -\frac{1}{2}$. This cumulative function is also an Incomplete Beta function found by setting $x = 2(1 - R')/3$.

$N = 4L (p = 4)$

If $N = 4L$, $\lambda_k'' = \cos \frac{2\pi f''}{N} = +1, 0, -1$. The fitted terms in m_i corresponding to $\lambda'' = 1$ have $f'' = 4m$, those for $\lambda'' = -1$ have $f'' = 4m - 2$; and those for $\lambda'' = 0$ have $f'' = 4m - 1$ or $4m - 3$. Let the number of terms in m_i of each sort be $L - n_1$, $L - n_2$, and $2L - n_3$, respectively. Then

$$D(R) = c \begin{cases} (1 + R)^{\frac{1}{2}(n_1+n_3-2)} \displaystyle\int_{y=0}^{1} y^{\frac{1}{2}(n_3-2)} (1 - y)^{\frac{1}{2}(n_1-2)} \\ \qquad \cdot [(1 - R) - y(1 + R)]^{\frac{1}{2}(n_2-2)}\, dy, & \text{for } R \leq 0, \\ (1 - R)^{\frac{1}{2}(n_2+n_3-2)} \displaystyle\int_{y=0}^{1} y^{\frac{1}{2}(n_3-2)}(1 - y)^{\frac{1}{2}(n_2-2)} \\ \qquad \cdot [(1 + R) - y(1 - R)]^{\frac{1}{2}(n_1-2)}\, dy, & \text{for } R \geq 0, \end{cases} \tag{54}$$

where R is ${}_LR_4$ and $c = \Gamma(\frac{1}{2}[n_1 + n_2 + n_3])/[\Gamma(\frac{1}{2}n_1)\Gamma(\frac{1}{2}n_2)\Gamma(\frac{1}{2}n_3)2^{\frac{1}{2}(n_1+n_2-2)}]$.

5.4. *The exact distribution of ${}_LR$ when $\rho \neq 0$.* The joint distribution of the observations for lag 1 when the null hypothesis is not true ($\rho \neq 0$) is (15), where Q is given by (45) with $L = 1$ and ${}_1C = RV$. V, R, $\{c_j\} (j = 1, \cdots, K')$ are a sufficient set of statistics for estimating σ^2, ρ, and $\{\gamma_j\} (j = 1, \cdots, K')$. Using the results given by Madow [11], it can be shown that the simultaneous distribution of V and R is

$$\frac{1 - \rho^N}{2^{\frac{1}{2}K''}\Gamma(\frac{1}{2}K'') \sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{1j'})}} V^{\frac{1}{2}K''-1} e^{-V(1+\rho^2-2\rho R)/2\sigma^2} D(R), \tag{55}$$

where $D(R)$ is the density function corresponding to (48). Integrating V from 0 to ∞, we obtain as the density for R

$$\frac{(1 - \rho^N)(\frac{1}{2}K'' - 1)}{\sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{1j'})}} (1 + \rho^2 - 2\rho R)^{\frac{1}{2}K''} \cdot \sum_{k=1}^{m} (-1)^{k+1}(\lambda_k'' - R)^{\frac{1}{2}(K''-4)} \mid V_k \mid, \tag{56}$$

for $\lambda''_{m+1} \leq R \leq \lambda''_m$, where V_k are given by (50). In the same way, one obtains the distribution of ${}_LR$ for $\rho \neq 0$ when $N = 2L$, $N = 3L$, and $N = 4L$ by multiplying (52), (53), and (54), respectively, by

$$(57) \qquad (1 - \rho^{\nu})^L \frac{(1 + \rho^2 - 2\rho R)^{\frac{1}{2}K''}}{\sqrt{\prod_{j'=1}^{K'} (1 + \rho^2 - 2\rho\lambda_{L,j'})}},$$

where $K'' = n_1 + n_2$ or $n_1 + n_2 + n_3$. This method was used by Madow for residuals from the sample mean [12].

6. Moments.

6.1. *The exact moments of R.* Most of the results of this section are straightforward adaptations of earlier results for the case of μ_i constant. Hence, we shall omit the details of derivations. The moment generating function of V and C for $\sigma^2 = 1$ is

$$(58) \qquad \phi(t_0, t) = E(e^{t_0 V + tC}) = \frac{1 - \rho^N}{\prod_{j''=K'+1}^{N} \left[1 + \rho^2 - 2t_0 - 2(\rho + t)\lambda_{1,j''}\right]^{\frac{1}{2}}}.$$

The h^{th} moment of $R = C/V$ is given by

$$(59) \qquad \mu'_h(R) = \int_{-\infty}^{0} \int_{-\infty}^{y_{h-1}} \cdots \int_{-\infty}^{y_1} \left.\frac{\partial^h \phi}{\partial t^h}\right]_{t=0} dt_0 \prod_{i=1}^{h-1} dy_i,$$

with the $\{y_i\}$ restricted from being too large (not more than a certain amount larger than zero). In the case of independence, ($\rho = 0$), we have the following first two moments of R:

$$(60) \qquad \begin{aligned} \mu'_1(R) &= \frac{1}{K''} \sum_{j''=K'+1}^{N} \lambda_{1,j''}; \\ \mu'_2(R) &= \frac{2}{K''(K''+2)} \sum_{j''=K'+1}^{N} \lambda_{1,j''}^2 + \frac{K''}{K''+2} [\mu'_1(R)]^2. \end{aligned}$$

If the $\lambda_{1,j''}$ are symmetrical (i.e. for each $\lambda_{1,j''}$, there is a $\lambda_{1,k''} = -\lambda_{1,j''}$), the mean of R is 0. For example, if 1 and $(-1)^i$ are fitted for N even, the mean is 0.

6.2. *Approximate moments of R when* $\rho = 0$. Since R and V are independent [8] when $\rho = 0$, $\mu'(R) = \mu'(C)/\mu'(V)$. V is a sum of squares and its moments are the same as for χ^2 with $N - K' = K''$ degrees of freedom. Using methods similar to those given by Dixon [5], we see that the moment generating function for C is

$$(61) \qquad \phi(t) = \alpha(t) \cdot \beta(t) \cdot \gamma(t),$$

where

$$(62) \qquad \begin{aligned} &\alpha(t) = \left(\frac{2}{A}\right)^{\frac{1}{2}N}, \beta(t) = A^N/[A^N - (2t)^N], \\ &\gamma(t) = \textstyle\prod_{j'} (1 - 2t\,\lambda_{1,j'}), \text{ and } A = 1 + \sqrt{1 - 4t^2}. \end{aligned}$$

In this case, $\lambda_{1,j'} = \cos \frac{2\pi j'}{N}$ includes all K' terms corresponding to those in m_i. Since the first N derivatives of $\beta(t)$ are zero at $t = 0$, we can use

$$\tilde{\phi}(t) = \alpha(t) \cdot \gamma(t) = \frac{2^{\frac{1}{2}N} \prod_{j'} (1 - 2t\,\lambda_{1,j'})^{\frac{1}{2}}}{(1 + \sqrt{1 - 4t^2})^{\frac{1}{2}N}} \tag{63}$$

as an approximation to (61). This expression yields the exact moments of C up to order N.

As a special case, consider $K' = 3$, with $\lambda_{1,1} = 1$ and $\lambda_{1,2} = \lambda_{1,3} = \cos \frac{2\pi j^*}{N}$. In this case

$$\tilde{\phi}_3(t) = \left(1 - 2t \cos \frac{2\pi j^*}{N}\right) \tilde{\phi}_1(t). \tag{64}$$

Successive derivatives of (64) at $t = 0$ show that

$$\mu_h'(R_3) = \left[P\,\mu_h'(R_1) - 2hQ \cos \frac{2\pi j^*}{N}\,\mu_{h-1}'(R_1)\right], \tag{65}$$

where $P = \mu_h'(V_1)/\mu_h'(V_3) = (N - 3 + 2h)/(N - 3)$, $Q = \mu_{h-1}'(V_1)/\mu_h'(V_3) = 2/(N - 3)$, and $h = 1, 2, \cdots, N$.

6.3. *Approximate moment generating function of C and V when $\rho \neq 0$.* To obtain an approximate moment generating function for C and V when $\rho \neq 0$, we utilize an approximation method given by Leipnik [10]. The exact moment generating function (58) with $\sigma^2 = 1$ can be written as

$$\phi(t_0, t) = (1 - \rho^N)\theta \exp. \left\{-\tfrac{1}{2} \sum_{i=1}^{N} \log\left[1 + \rho^2 - 2t_0 - 2(\rho + t) \cos \frac{2\pi i}{N}\right]\right\}, \tag{66}$$

where $\theta = \prod_{j'}[1 + \rho^2 - 2t_0 - 2(\rho + t)\lambda_{1j'}]^{\frac{1}{2}}$, and j' refers to the K' fitted terms in m_i. If the sum in the exponent of (66) is replaced by

$$\int_0^N \log\left[1 + \rho^2 - 2t_0 - 2(\rho + t) \cos \frac{2\pi x}{N}\right] dx, \tag{67}$$

and if $(1 - \rho^N)$ is replaced by 1, we obtain the approximate moment generating function

$$\tilde{\phi} = \frac{\prod_{j'} [1 + \rho^2 - 2t_0 - 2(\rho + t)\,\lambda_{1,j'}]^{\frac{1}{2}}}{[\frac{1}{2}(1 + \rho^2 - 2t_0 + \sqrt{(1 + \rho^2 - 2t_0)^2 - 4(\rho + t)^2}\,)]^{\frac{1}{2}N}}. \tag{68}$$

7. Approximate distributions of R.

7.1. *The Pearson Type* I (*Incomplete Beta*) *distribution*. The significance points of ${}_LR$ can be found exactly from equation (48) for $L = 1$ and by integrating equations (52), (53), and (54) for $N = 2L$, $3L$, and $4L$, respectively. These exact probability integrals for $N = 2L$, $3L$, and $4L$ are simply sums of Incomplete Beta functions, and the significance points can be found in Pearson's *Tables of*

the Incomplete Beta-Function [14] or in the Thompson tables [16]. However, the computation of the exact significance points for $L = 1$ and $N > 4$ by use of equation (48) is quite tedious and actually impossible for large N with present logarithm tables and readily available computing devices. Hence, approximate distributions are called for.

The Type I approximation to the distribution of R is

$$f_1(R) = \frac{(1 + R)^{p-1} (1 - R)^{q-1}}{2^{p+q-1} \beta(p, q)}, \; -1 \leq R \leq 1, \tag{69}$$

where p and q are chosen so that the first two moments of this approximate distribution agree with the first two moments of the exact distribution. It can be shown that each moment of the approximate distribution approaches the corresponding exact moment quite rapidly as N increases. On the basis of the approximation, the probability α of the significance point R' being exceeded can be found from the Incomplete Beta function. Thus

$$\alpha = Pr\{R > R'\} = 1 - I_x(p, q) = I_{x'}(p', q'), \tag{70}$$

where

$$I_x(p, q) = \frac{1}{\beta(p, q)} \int_0^x y^{p-1} (1 - y)^{q-1} \, dy, \tag{71}$$

and $x = (1 + R')/2$, $x' = (1 - x)$, $p' = q$, and $q' = p$. Hence, $R' = 2x - 1 = 1 - 2x'$.

The parameters in (69) are taken to be

$$2p = (1 + \mu_1')(1 - \mu_2')/\mu_2, \qquad 2q = (1 - \mu_1')(1 - \mu_2')/\mu_2, \tag{72}$$

where $\mu_2 = \mu_2' - (\mu_1')^2$ and $\mu_i' - \mu_i'(R)$ given in (60). Hence, when the distribution of R is symmetric, $\mu_1' = 0$ and $2p = 2q = (1 - \mu_2')/\mu_2'$.

In Section 3.1, we set up significance points for four special trends for which $\mu_1' = 0$:

(b) $P = 2$; (c) $P = 2, 4$; (d) $P = 2, 3, 6$; (e) $P = 2, 12/5, 3, 4, 6, 12$.
The values of μ_2' for these four trends are:
(b) $(N - 4)/[N(N - 2)]$, (c) $1/(N - 2)$, (d) $1/(N - 4)$, (e) $1/(N - 10)$.
Naturally the third moments for these symmetric distributions are 0. The fourth moments are as follows:

Trend	(b)	(c)	(d)	(e)
Exact	$\frac{3(N^2 - 2N - 16)}{(N+4)(N+2)N(N-2)}$	$\frac{3(N^2 - 2N - 16)}{(N+2)(N)(N-2)(N-4)}$	$\frac{3}{(N-2)(N-4)}$	$\frac{3}{(N-8)(N-10)}$
Incomplete Beta	$\frac{3(N-4)^2}{N(N-2)(N^2-8)}$	$\frac{3}{N(N-2)}$	$\frac{3}{(N-2)(N-4)}$	$\frac{3}{(N-8)(N-10)}$

We note that for (d) and (e), the fourth moments for the Incomplete Beta are exact and for (b) and (c), they approach the exact values quite rapidly as N increases.

In Section 3.2, we considered some significance points for the following single-period trends: $P = 3, 4, 6$, and 12. The values of $2p$ and $2q$ for these asymmetrical cases are

$$2p = \frac{(N - 4 - 2\lambda)E}{D}; \qquad 2q = \frac{(N - 2 + 2\lambda)E}{D}, \tag{73}$$

where $\lambda = \cos \frac{2\pi}{P}$, $E = (N - 1)(N - 4) - 4\lambda$ and $D = (N - 3)(N - 1 + 4\lambda) - (N - 1)(1 + 2\lambda)^2$.

Equation (69) has the drawback of using the range $(-1, +1)$ instead of the true range of R, which varies between the last (smallest) λ_k'' to the first (largest) λ_k''. For example, if $N = 12$ and we fit the constant, $\cos \frac{2\pi i}{12}$, and $\sin \frac{2\pi i}{12}$, then $\lambda_{1,1} = 1$, $\lambda_{1,2} = \lambda_{1,3} = \cos \frac{2\pi}{12} = \frac{\sqrt{3}}{2}$, and the range of R is $\left(-1, \cos \frac{4\pi}{12} = \frac{1}{2}\right)$. However, if we fit the constant and $\cos \pi i = (-1)^i$, then $\lambda_{1,1} = 1$ and $\lambda_{1,2} = -1$, the true range would be $\left(-\frac{\sqrt{3}}{2}, +\frac{\sqrt{3}}{2}\right)$. From these examples we see that the error in using the approximate range $(-1, +1)$ varies according to the fitted terms in m_i, and that the error is worse on one tail than on the other, unless symmetric terms are fitted. A more accurate approximation could be obtained by use of the exact curtailed range, but it was not thought desirable because the exact range rapidly approaches the approximate range as N increases.

We might add that the significance point, R', can also be calculated from the Inverted Beta (F) distribution, for which tables are given by Merrington and Thompson [13], Snedecor [15], and Fisher and Yates [6]. Cochran [4] has provided an approximate formula for $z = \frac{1}{2} \log_e F$ when n_1 and n_2 are not given in the F-tables.

7.2. *The normal approximation.* It should be noted that R is asymptotically normally distributed for $\rho = 0$, as shown by the form of the characteristic function. We have considered the normal approximation with mean $\mu_1'(R)$ and variance $\mu_2(R)$. The variance of R was given in the previous section for the four special trends. For all single period trends, except $P = 2$, $\mu_1' = -(1 + 2\lambda)/(N - 3)$ and the variance is

$$\mu_2 = \frac{(N - 1 + 4\lambda)}{(N-1)(N-3)} - (\mu_1')^2, \tag{74}$$

where, as before, $\lambda = \cos(2\pi/P)$. Further terms in an asymptotic expansion of the distribution would take account of higher moments of R as Hsu has done for the case of fitting only the mean (m_i = a constant) [7].

REFERENCES

[1] R. L. Anderson, *Serial correlation in the analysis of time series*, unpublished thesis, Library, Iowa State College, 1941.

[2] R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 1-13.

[3] T. W. ANDERSON, "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, Vol. 31 (1948), pp. 88–116.

[4] W. G. COCHRAN, "Note on an approximate formula for the significance levels of z," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 93–95.

[5] W. J. DIXON, "Further contributions to the problem of serial correlation," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[6] R. A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural, and Medical Research*, 2d ed., Oliver and Boyd Ltd., 1943.

[7] P. L. HSU, "On the asymptotic distributions of certain statistics used in testing the independence between successive observations from a normal population," *Annals of Math. Stat.*, Vol 17 (1946), pp. 350–354.

[8] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 14–23.

[9] E. L. LEHMANN, "On optimum tests of composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 18 (1947), p. 481.

[10] R. B. LEIPNIK, "Distribution of the serial correlation coefficient in a circularly correlated universe," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 80–87.

[11] W. G. MADOW, "Contributions to the theory of multivariate statistical analysis," *Trans. Am. Math. Soc.*, Vol. 44 (1938), p. 461.

[12] W. G. MADOW, "Note on the distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 308–310.

[13] M. MERRINGTON AND C. M. THOMPSON, "Tables of percentage points of the inverted Beta (F) distribution," *Biometrika*, Vol. 13 (1943), pp. 73–88.

[14] K. PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge Univ. Press, 1934.

[15] G. W. SNEDECOR, *Statistical Methods*, 4th ed., Iowa State College Press, 1946, pp. 222–225.

[16] C. M. THOMPSON, "Tables of percentage points of the incomplete beta-function," *Biometrika*, Vol. 32 (1941), pp. 151–181.

[17] UNITED STATES DEPARTMENT OF AGRICULTURE, *Agricultural Statistics*, United States Government Printing Office, Washington, D. C., 1939, p. 390.

BAYES SOLUTIONS OF SEQUENTIAL DECISION PROBLEMS

By A. Wald and J. Wolfowitz

Columbia University

Summary. The study of sequential decision functions was initiated by one of the authors in [1]. Making use of the ideas of this theory the authors succeeded in [4] in proving the optimum character of the sequential probability ratio test. In the present paper the authors continue the study of sequential decision functions, as follows:

a) The proof of the optimum character of the sequential probability ratio test was based on a certain property of Bayes solutions for sequential decisions between two alternatives, the cost function being linear. This fundamental property, the convexity of certain important sets of a priori distributions, is proved in Theorem 3.9 in considerable generality. The number of possible decisions may be infinite.

b) Theorem 3.10 and section 4 discuss tangents and boundary points of these sets of a priori distributions.

(These results for finitely many alternatives were announced by one of us in an invited address at the Berkeley meeting of the Institute of Mathematical Statistics in June, 1948)[1]

c) Theorem 3.6 is an existence theorem for Bayes solutions. Theorem 3.7 gives a necessary and sufficient condition for a Bayes solution. These theorems generalize and follow the ideas of Lemma 1 of [4]

d) Theorems 3.8 and 3.8.1 are continuity theorems for the average risk function. They generalize Lemma 3 in [4]

e) Other theorems give recursion formulas and inequalities which govern Bayes solutions.

1. Introduction. In a previous publication of one of the authors [1] the decision problem was formulated as follows: Let $X = \{x_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of chance variables. An observation on X is given by a sequence $x = \{x_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ of real values, where x_i denotes the observed value of X_i. A sequence x is also called a sample or sample point, and the totality M of all possible sample points x is called the sample space. Let $G(x)$ denote the probability that $X_i < x_i$ for $i = 1, 2, \cdots, \text{ad inf.}$; i.e., G is the cumulative distribution function of X. In a statistical decision problem G is assumed to be unknown. It is merely known that G is an element of a given class Ω of distribution functions. There is given, furthermore, a space D^* whose elements d represent the possible decisions that can be made in the problem under consideration.

[1] A brief statement of some of the results of the present paper is to be found in the authors' paper of the same name in the *Proc. Nat. Acad. Sci. U. S. A.*, Vol. 35 (1949), pp. 99-102.

The problem is to construct a function $d = D(x)$, called **the decision function,** which associates with each sample point x an element d of D^* so that the decision $d = D(x)$ is made when x is observed.

Occasionally we shall use the symbol D to denote a decision function $D(x)$. This will be done especially when we want to emphasize that we mean the whole decision function and not merely a particular value of it corresponding to some particular x.

If $d = D(x)$ is the decision function adopted and if $x^0 = \{x_i^0\}$ $(i = 1, 2, \cdots)$ is the particular sample point observed, the number of components of x^0 we have to observe in order to reach a decision is equal to the smallest positive integer $n = n(x^0)$ with the property that $D(x) = D(x^0)$ for any x for which $x_1 = x_1^0, \cdots, x_n = x_n^0$. If no finite n exists with the above property, we put $n(x) = \infty$. If $d(x)$ is equal to a constant d, we put $n(x) = 0$. We shall call $n(x)$ the number of observations required by D when x is the observed sample. Of course, $n(x)$ depends also on the decision rule D adopted. To put this in evidence, we shall occasionally write $n(x, D)$ instead of $n(x)$. If D_0 is a decision function such that $n(x, D_0)$ has a constant value over the whole sample space M, we have the classical non-sequential case. If $n(x, D_0)$ is not constant, we shall say that D_0 is a sequential decision function.

In the remainder of this section we shall sketch briefly some of the fundamental notions of the theory without regard to regularity conditions. The latter will be discussed in the next section.

In [1] a weight function $W(G, d)$ was introduced which expresses the loss suffered by the statistician when G is the true distribution of X and the decision d is made. Let $c(n)$ denote the cost of making n observations; i.e., $c(n)$ is the cost of observing the values of $X_1, \cdots, X_n$. Then, if the decision function $d = D(x)$ is adopted and G is the true distribution of X, the expected value of the loss due to possible erroneous decisions plus the expected cost of experimentation is given by

$$r(G, D) = \int_M W[G, D(x)]\, dG(x) + \int_M c[n(x, D)]\, dG(x). \tag{1.1}$$

The above expression is called the risk when D is the decision function adopted and G is the true distribution.

Let ξ be an a priori probability distribution on Ω; i.e., ξ is a probability measure defined over a suitably chosen Borel field[2] of subsets of Ω. Then the expected value of $r(G, D)$ is given by

$$r(\xi, D) = \int_\Omega r(G, D)\, d\xi. \tag{1.2}$$

[2] A Borel field is an aggregate of sets such that a) the null set is a member of the field, b) the complement with respect to the entire space (here M) is a member of the field, c) the sum of denumerably many members of the field is itself in the field.

The above expression is called the risk when ξ is the a priori distribution on Ω and D is the decision function adopted.

We shall say that the decision function D_0 is a Bayes solution relative to the a priori distribution ξ if

$$(1.3) \qquad r(\xi, D_0) \leqq r(\xi, D) \text{ for all } D.$$

If there existed an a priori distribution on Ω and if this distribution were known, we could put ξ equal to this a priori distribution and a Bayes solution relative to ξ would provide a very satisfactory solution of the decision problem. In most applications, however, not even the existence of an a priori distribution can be postulated. Nevertheless, the study of Bayes solutions corresponding to various a priori distributions is of great interest in view of some results given in [1]. It was shown in [1] that under rather general conditions the class C of the Bayes solutions corresponding to all possible a priori distributions ξ has the following property: If D_1 is a decision function that is not an element of C, there exists a decision function D_2 in C such that

$$(1.4) \qquad r(G, D_2) \leqq r(G, D_1) \text{ for all } G$$

and

$$(1.5) \qquad r(G, D_2) < r(G, D_1) \text{ for at least one } G.$$

It was furthermore shown in [1] that under general conditions a minimax solution D_0 of the decision problem is also a Bayes solution corresponding to some a priori distribution ξ. By a minimax solution we mean a decision function D_0 such that, for all D

$$(1.6) \qquad \operatorname*{Sup}_G r(G, D_0) \leqq \operatorname*{Sup}_G r(G, D).$$

2. Regularity conditions and other assumptions. We shall make the following assumptions:

Assumption 1. *The chance variables are identically and independently distributed. The common distribution is either discrete or absolutely continuous.*

Let $p(a \mid F)$ denote the elementary probability law of X_i when F is the distribution of X_i; i.e., when F is discrete, $p(a \mid F)$ is the probability that $X_i = a$, and when F is absolutely continuous, $p(a \mid F)$ is the probability density of X_i at a.

In the space M of sequences x let B be the smallest Borel field which contains all sets of points x which are defined by the relations

$$x_i < a_i \qquad i = 1, 2, \cdots \text{ ad inf.},$$

where the a_i are real numbers or $+\infty$. Each admissible[3] F induces a probability measure $F^*(B)$ on M; the totality of these probability measures is Ω. Let H^*

[3] An F or F^* is admissible if F^* is in Ω.

be a given Borel field of subsets of Ω. The only subsets of Ω which we shall discuss in this paper will be members of H^*, and all probability measures on Ω which we shall discuss will be measurable (H^*). This will henceforth be assumed without further repetition.

Let A^* be any set in H^*, and A the set of F which corresponds to the F^* in A^*. The sets A form a Borel field, say H. By definition, the probability measure of a set A according to a probability measure $\xi(H^*)$ on Ω is to be the same as the probability measure of A^* according to ξ.

Let $M \times \Omega$ be the Cartesian product of M and Ω ([5], page 82), and K be the smallest Borel field of subsets of $M \times \Omega$ which contains the Cartesian product of any member of B by any member of H^*.

For a given decision function $d = D(x)$, $W(F, D(x))$ is a function of F and x. Hereafter in this paper we shall limit ourselves to functions $D(x)$ such that $W(F, D(x))$ is measurable (K), and $n(x, D)$ is measurable (B).

It is true that in Section 1, W was given as a function of G, the distribution of X. Because of Assumption 1, $G = F^*$, and there is a one-to-one correspondence between F and F^*. Thus we may, in appropriate places, interchange them freely.

Assumption 2. *For every real a, except possibly on a Borel set[4] whose probability is zero according to every admissible F, $p(a \mid F)$ exists and is a function of a and F which is measurable (K). If the admissible distributions F are discrete, there exists a fixed sequence $\{b_i\}$ ($i = 1, 2, \cdots$, ad inf.) of real values such that $\sum_{i=1}^{\infty} p(b_i \mid F) = 1$ for all admissible F.*

Assumption 3. *$W(F, d)$ is bounded. For every d in D^*, $W(F, d)$ is a function of F which is measurable (H).*

In what follows ξ will always denote a probability measure (H^*) on Ω. Thus

$$W(\xi, d) = \int_{\Omega} W(F, d)\, d\xi$$

exists.

Assumption 4. *The function $c(n) = cn$. Without loss of generality we may take $c = 1$, so that $c(n) = n$.*

We shall introduce the following convergence definition in the space D^*: the sequence $\{d_i\}$ converges to d_0 if

$$\lim_{i \to \infty} W(F, d_i) = W(F, d_0)$$

uniformly in the admissible F's.

Assumption 5. *The space D^* is compact in the sense of the above convergence definition.*

One can easily verify that, if $\lim_{i \to \infty} d_i = d_0$, then

$$\lim_{i \to \infty} W(\xi, d_i) = W(\xi, d_0);$$

[4] A Borel set is a member of the smallest Borel field which contains all the open sets of the real line.

i.e., $W(\xi, d)$ is a continuous function of d. Thus, because of Assumption 5, the minimum of $W(\xi, d)$ with respect to d exists.

We shall now show that, under the above conditions

$$\int_M W[F^*, D(x)]\, dF^*(x) \tag{2.1}$$

exists and is a function of F^* measurable (H^*). For any j let R_j be the set in B such that $n(x, D) = j$. Then it is enough to show that, for any j,

$$\int_{R_j} W[F^*, D(x)]\, dF^*(x) \tag{2.2}$$

exists and is a function of F^* measurable (H^*).

In the discrete case, the integral (2.2) is equal to the sum[5]

$$\sum_{(x_1,\dots,x_j)\epsilon R_j} W[F^*, D(x)]p(x_1 \mid F) \cdots p(x_j \mid F). \tag{2.3}$$

For fixed values of $x_1, \cdots, x_j$, the expression under the summation sign is obviously a function of F^* measurable (H^*). Since, because of Assumption 2, there are only countably many points $(x_1, \cdots, x_j)$ in R_j, the sum (2.3) must be a function of F^* measurable (H^*).

In the absolutely continuous case, the integral (2.2) is equal to (2.4)

$$\int_{R_j} W[F^*, D(x)] \prod_{i=1}^{j} p(x_i \mid F)\, d\nu(j) \tag{2.4}$$

where $\nu(j)$ is Borel measure in the j-dimensional Euclidean space. The integrand is measurable (K). Hence, the integral (2.4) exists and is a function of F^* measurable (H^*) (see [5], Chapter III, Theorems 9.3 and 9.8).

3. Some results concerning Bayes solutions. If ξ is the a priori probability measure on Ω, the a posteriori probability of a subset ω of Ω for given values $x_1, \cdots, x_m$ of the first m chance variables is given by

$$\xi(\omega \mid \xi, x_1, \cdots, x_m) = \frac{\int_\omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}{\int_\Omega p(x_1 \mid F) \cdots p(x_m \mid F)\, d\xi}. \tag{3.1}$$

Let

$$\rho_0(\xi) = \operatorname*{Min}_d W(\xi, d). \tag{3.2}$$

For any positive integral value m, let $\rho_m(\xi)$ denote the infimum of $r(\xi, D)$ with respect to D where D is restricted to decision functions for which $n(x, D) \leqq m$ for all x. For any positive integer m, let $d = D^m(x)$ denote a decision function

[5] Because of the definition of R_j we may, in the expressions (2.3) and (2.4), proceed as if R_j were a Borel set in j-dimensional Euclidean space.

D for which $n(x, D) \leqq m$ for all x. Thus, we can write

$$\rho_m(\xi) = \operatorname*{Inf}_{D^m} r(\xi, D^m) \quad (m = 1, 2, \cdots, \text{ad inf.}). \tag{3.3}$$

Let

$$\rho(\xi) = \operatorname*{Inf}_{D} r(\xi, D). \tag{3.4}$$

We shall first prove several theorems concerning the functions $\rho_0(\xi)$, $\rho_m(\xi)$, and $\rho(\xi)$.

THEOREM 3.1. *The following recursion formula holds:*[6]

$$\rho_{m+1}(\xi) = \operatorname{Min}\left[\rho_0(\xi), 1 + \int_{-\infty}^{\infty} \rho_m(\xi_a)\, p(a \mid \xi)\, da\right] \tag{3.5}$$

$$(m = 0, 1, 2, \cdots, \text{ad inf.})$$

where

$$\xi_a(\omega) = \xi(\omega \mid \xi, a) \text{ and } p(a \mid \xi) = \int_\Omega p(a \mid F)\, d\xi. \tag{3.6}$$

PROOF: Let $\rho_m^*(\xi)$ $(m = 1, 2, \cdots, \text{ad inf.})$ denote the infimum of $r(\xi, D)$ with respect to D where D is subject to the restriction that $n(x, D) \geqq 1$ and $\leqq m$ for all x. Clearly,

$$\rho_{m+1}(\xi) = \operatorname{Min}[\rho_0(\xi), \rho_{m+1}^*(\xi)]. \tag{3.7}$$

Let $\rho_m^*(\xi \mid a)$ denote the infimum with respect to D of the conditional risk (conditional expected value of $W[F, D(x)] + n(x, D)$) when the first observation x_1 on X_1 is a and D is restricted to decision functions for which $n(x, D) \geqq 1$ and $\leqq m$ for all x. Let $\bar{D}(m)$ be the temporary generic designation of such a decision function. Let $\bar{D}(m \mid a)$ be the decision function which is obtained from $\bar{D}(m)$ when the first observation is a. Finally let $r(\xi, D \mid a)$ be the conditional risk when the a priori distribution function is ξ, D is the decision function and requires at least one observation, and the first observation is a. We then have that

$$r(\xi, \bar{D}(m + 1) \mid a) = r(\xi_a, \bar{D}(m + 1 \mid a)) + 1.$$

Hence

$$\rho_{m+1}^*(\xi \mid a) = \rho_m(\xi_a) + 1. \tag{3.8}$$

The unconditional quantity $\rho_{m+1}^*(\xi)$ must clearly be equal to the average value of the infimum of the conditional risk. Thus we have

$$\rho_{m+1}^*(\xi) = \int_{-\infty}^{\infty} \rho_{m+1}^*(\xi \mid a) p(a \mid \xi) da. \tag{3.9}$$

---

[6] If the distribution of X is discrete, the integration with respect to a is to be replaced by summation with respect to a. This remark refers also to subsequent formulas.

Equation (3.5) follows from (3.7), (3.8) and (3.9).

THEOREM 3.2. *The function* $\rho(\xi)$ *satisfies the following equation:*

$$\rho(\xi) = \text{Min}\left[\rho_0(\xi), \int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da + 1\right]. \tag{3.10}$$

The proof of this theorem is omitted, since it is essentially the same as that of Theorem 3.1.

THEOREM 3.3.[7] *The following inequalities hold:*

$$0 \leqq \rho_m(\xi) - \rho(\xi) \leqq \frac{W_0^2}{m} \qquad (m = 1, 2, \cdots, \text{ad inf.}) \tag{3.11}$$

where W_0 *is the least upper bound of* $W(F, d)$.

PROOF: Let $\{D_i\}$ $(i = 1, 2, \cdots, \text{ad inf.})$ be a sequence of decision functions such that

$$\lim_{i=\infty} r(\xi, D_i) = \rho(\xi). \tag{3.12}$$

Let, furthermore, $P_i(\xi)$ denote the probability that at least m observations will be made when D_i is the decision function adopted and ξ is the a priori probability measure on Ω. Since $\rho(\xi) \leqq W_0$ and since

$$r(\xi, D_i) \geqq mP_i(\xi), \tag{3.13}$$

it follows from (3.12) that

$$\limsup_{i=\infty} P_i(\xi) \leqq \frac{W_0}{m}. \tag{3.14}$$

Let D_i^m be the decision function obtained from D_i as follows: $D_i^m(x) = D_i(x)$ for all x for which $n(x, D_i) \leqq m$. $D_i^m(x)$ is equal to a fixed element d_0 for all x for which $n(x, D_i) > m$.[8]

Clearly,

$$r(\xi, D_i^m) \leqq r(\xi, D_i) + P_i(\xi)W_0. \tag{3.15}$$

From (3.12), (3.14) and (3.15) it follows that

$$\limsup_{i=\infty} r(\xi, D_i^m) \leqq \rho(\xi) + \frac{W_0^2}{m}. \tag{3.16}$$

Since $\rho_m(\xi)$ cannot exceed the left hand member of (3.16), the second half of (3.11) follows from (3.16). The first half of (3.11) is obvious.

---

[7] This theorem is essentially the same as Lemma 2.1 in [6].

[8] We verify that $W(F, D_i^m)$ is measurable (K), as follows: Consider the set V of couples (F, x) such that $W(F, D_i^m(x)) < c$, where c is some real constant. We want to show that $V \epsilon K$. For this purpose let V_0 be the set of couples (F, x) such that $W(F, D_i(x)) < c$. Then $V_0 \epsilon K$. Let V_1 be the set of x's such that $n(x, D_i) \leqq m$. Then $V_1 \epsilon B$, $(\Omega \times V_1) = V_2 \epsilon K$, $V_0V_2 \epsilon K$. Let $V_3 = M - V_1$. For every $x \epsilon V_3$ we have $W(F, D_i^m(x)) = W(F, d_0)$. Let V_4 be the set of F's such that $W(F, d_0) < c$. Then $V_4 \epsilon H$ by Assumption 3. Finally we have $V = V_0V_2 + V_4 \times V_3$, so that $V \epsilon K$.

The immediate consequence of Theorem 3.3 is the relation[9]

$$\lim_{m=\infty} \rho_m(\xi) = \rho(\xi). \tag{3.17}$$

THEOREM 3.4. *If ξ_1 and ξ_2 are two probability measures on Ω such that*[10]

$$\frac{\xi_1(\omega)}{\xi_2(\omega)} \leqq 1 + \epsilon \text{ for all } \omega, \tag{3.18}$$

then

$$\rho(\xi_1) \leqq (1 + \epsilon)\rho(\xi_2). \tag{3.19}$$

PROOF: It follows from (3.18) that

$$r(\xi_1 , D) \leqq (1 + \epsilon)r(\xi_2 , D) \text{ for all } D. \tag{3.20}$$

Hence, (3.19) must hold.

The above theorem permits the computation of a simple and in many cases useful lower bound of $\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da$ as follows:
For any real value a, let ϵ_a be a non-negative value (not necessarily finite) determined such that

$$\frac{\xi(\omega)}{\xi_a(\omega)} \leqq 1 + \epsilon_a \text{ for all } \omega. \tag{3.21}$$

Then

$$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \int_{-\infty}^{\infty} \frac{\rho(\xi)}{1 + \epsilon_a}\, p(a \mid \xi)\, da = \rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da. \tag{3.22}$$

Since $\epsilon_a \geqq 0$ and since $\rho_0(\xi) \geqq \rho(\xi)$, we obviously have

$$\rho(\xi) \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da \geqq \rho(\xi) - \left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right]\rho_0(\xi). \tag{3.23}$$

Hence, we obtain the inequality

$$\int_{-\infty}^{\infty} \rho(\xi_a)\, p(a \mid \xi)\, da \geqq \rho(\xi) - \rho_0(\xi)\left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da\right]. \tag{3.24}$$

An upper bound of the left hand member in (3.24) is obtained by replacing ρ by ρ_0 ; i.e.,

$$\int_{-\infty}^{\infty} \rho(\xi_a)p(a \mid \xi)\, da \leqq \int_{-\infty}^{\infty} \rho_0(\xi_a)p(a \mid \xi)\, da. \tag{3.25}$$

---

[9] A proof of (3.17) is contained implicitly in the work of Arrow, Blackwell and Girshick ([2], Section 1.3).

[10] The left member of (3.18) is defined to be equal to 1 when $\xi_1(\omega) = \xi_2(\omega) = 0$.

The bounds given in (3.24) and (3.25) may be useful in constructing Bayes solutions, since the following theorem holds:

THEOREM 3.5. *If*

$$\rho_0(\xi) > \int_{-\infty}^{\infty} \rho_0(\xi_a) p(a \mid \xi)\, da + 1, \tag{3.26}$$

then $\rho(\xi) < \rho_0(\xi)$. *If*

$$\rho_0(\xi) \left[1 - \int_{-\infty}^{\infty} \frac{p(a \mid \xi)}{1 + \epsilon_a}\, da \right] < 1, \tag{3.27}$$

then $\rho(\xi) = \rho_0(\xi)$.

The above theorem is an immediate consequence of (3.10), (3.24) and (3.25).

A decision procedure relative to a given a priori probability measure ξ_0 will be given with the help of the function $\rho(\xi)$ as follows: If $\rho(\xi_0) = \rho_0(\xi_0)$, take a final decision d for which $W(\xi_0, d)$ is minimized. If $\rho(\xi_0) < \rho_0(\xi_0)$, take an observation on X_1 and compute the a posteriori probability measure ξ_1. If $\rho(\xi_1) = \rho_0(\xi_1)$, stop experimentation with a final decision d for which $W(\xi_1, d)$ is minimized. If $\rho(\xi_1) < \rho_0(\xi_1)$, take an observation on X_2 and compute the a posteriori probability measure ξ_2 corresponding to the observed values of X_1 and X_2, and so on. The above decision procedure will be shown later to be a Bayes solution. Theorem 3.5 permits one to decide whether $\rho(\xi) < \rho_0(\xi)$ or $= \rho_0(\xi)$ whenever ξ satisfies (3.26) or (3.27). Theorem 3.5 will be useful when the class of all ξ's for which neither (3.26) nor (3.27) holds is small.

For the purposes of the next theorem let $\hat{D}$ designate the decision procedure described in the preceding paragraph. (We shall shortly show that $\hat{D}$ is a decision function in the sense of our definition.)

Let $\hat{D}^0$ be the decision procedure where the first observation is taken and then one proceeds according to $\hat{D}$.

We shall now prove that $\hat{D}$ and $\hat{D}^0$ are Bayes solutions. More precisely, we shall prove the following theorem:[11]

THEOREM 3.6. *For any* ξ, $\hat{D}$ *and* $\hat{D}^0$ *as defined above are decision functions. Let* D *be any decision function for which* $n(x, D) \geqq 1$ *and let*

$$\rho^*(\xi) = \operatorname*{Inf}_D r(\xi, D).$$

Then

$$r(\xi, \hat{D}) = \rho(\xi)$$

and

$$r(\xi, \hat{D}^0) = \rho^*(\xi).$$

[11] This theorem follows also from some earlier more general existence theorems ([6], Theorems 2.4 and 3.3). (See also [4], Lemma 1.) The validity of Theorem 3.6 was proved also by Arrow, Blackwell and Girshick [2].

In view of this theorem, the operation "infimum with respect to D" in the definitions of $\rho(\xi)$, and $\rho^*(\xi)$ can be replaced by "minimum with respect to D."

First we shall establish the measurability properties of $\hat{D}$ and $\hat{D}^0$. Since the proofs are similar, we restrict ourselves to consideration of $\hat{D}$. Let $\xi_{x_1,\dots,x_m}$ be the a posteriori distribution (3.1). From the (B) measurability of $\rho_0(\xi_{x_1,\dots,x_m})$ and $\rho(\xi_{x_1,\dots,x_m})$ it follows easily that $n(x, \hat{D})$ is measurable (B). It remains to prove that $W(F, \hat{D}(x))$ is measurable (K). For this purpose, let $L^i = (d_1^i, \cdots, d_{k_i}^i)$ be a sequence $\frac{1}{i}$ dense in D^*, i.e., for any $d \in D^*$ there exists a $g \in D^*$ such that $g \in L^i$ and $| W(F, d) - W(F, g) | < \frac{1}{i}$ uniformly in F. (The existence of such a sequence follows from Assumption 5.) Let now $D_i(x)$ be a decision function defined as follows:

$$n(x, D_i) = n(x, \hat{D}).$$

Suppose $n(x, \hat{D}) = m$ when the observations are $x_1, \cdots, x_m$. We define $D_i(x)$ to be such that $D_i(x)$ is an element of L^i and

$$W(\xi_{x_1,\dots,x_m}, D_i(x)) = \operatorname*{Min}_{d \in L^i} W(\xi_{x_1,\dots,x_m}, d), \tag{3.28}$$

i.e., $D_i(x)$ takes the minimizing value of d. For any fixed d, the set of x's satisfying the equation $D_i(x) = d$ is without difficulty shown to be (B) measurable. Since $D_i(x)$ assumes only a finite number of values in D^*, it follows from Assumption 3 that $W(F, D_i(x))$ is measurable (K). Now

$$\lim_{i \to \infty} W(F, D_i(x)) = W(F, \hat{D}(x)),$$

so that $W(F, \hat{D}(x))$ is measurable (K).

We shall now prove that $\hat{D}$ is a Bayes solution, i.e., that

$$\rho(\xi) = r(\xi, \hat{D}). \tag{3.29}$$

In a similar way it can be proved that

$$\rho^*(\xi) = r(\xi, \hat{D}^0). \tag{3.30}$$

If $\rho_0(\xi) = \rho(\xi)$, there can be no better decision function (from the point of view of reducing the risk) than $\hat{D}$, i.e., $\hat{D}$ is a Bayes solution. Suppose then that

$$\rho_0(\xi) > \rho(\xi). \tag{3.31}$$

If (3.31) holds and $\hat{D}$ is not a Bayes solution, there exists a decision function $\bar{D}_1$ such that

$$r(\xi, \bar{D}_1) < r(\xi, \hat{D}) \tag{3.32}$$

and

$$r(\xi, \bar{D}_1) < \frac{\rho_0(\xi) + \rho(\xi)}{2}. \tag{3.33}$$

Now $\bar{D}_1$ must require that at least one observation be taken, else (3.33) could not hold. Thus $\hat{D}$ and $\bar{D}_1$ both require at least one observation.

Suppose one observation is taken. Let $r(\xi, D \mid a)$ denote the conditional risk of proceeding according to D when ξ is the a priori distribution and a is the first observation. For a given D we have that $r(\xi, D \mid a)$ is a function only of ξ_a. In particular $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ are functions only of ξ_a.

We can now apply to $r(\xi, \hat{D} \mid a)$ and $r(\xi, \bar{D}_1 \mid a)$ the same argument that was applied above to $r(\xi, \hat{D})$ and $r(\xi, \bar{D}_1)$, and conclude again as follows: whenever $\rho_0(\xi_a) = \rho(\xi_a)$ (when one takes no more observations according to $\hat{D}$), taking additional observations cannot diminish the conditional risk below $r(\xi, \hat{D} \mid a)$ ($\bar{D}_1$ may require an additional observation without having

$$r(\xi, \bar{D}_1 \mid a) > r(\xi, \hat{D} \mid a).$$

This can happen when $\rho_0(\xi_a) = \rho^*(\xi_a)$). Whenever $\rho_0(\xi_a) > \rho(\xi_a)$ (when $\hat{D}$ requires us to take another observation) two cases may occur: either a) $\bar{D}_1$ requires us to take another observation, in which case its decision is the same as that of $\hat{D}$, or b) $\bar{D}_1$ requires us to stop taking observations. There exists then another decision function whose conditional risk is less than

$$\frac{\rho_0(\xi_a) + \rho(\xi_a)}{2} + 1.$$

Both this decision function and $\hat{D}$ require that another observation be taken. We conclude that up to and including the first observation, $\hat{D}$ coincides either with $\bar{D}_1$ or with another decision function $\bar{D}_2$ whose risk is not greater than that of $\bar{D}_1$.

We continue in this manner for 2, 3, $\cdots$ observations. The above argument is always valid because of Assumption 4 and because the past history of the process (the sequence of observations) enters only through the a posteriori probability. Thus we conclude that for any positive integer k there exists a decision function $\bar{D}_k$ such that up to and including the k-th observation $\hat{D}$ gives the same decision as $\bar{D}_k$ and the risk corresponding to $\bar{D}_k$ does not exceed the risk corresponding to $\bar{D}_1$. Since $\lim_{k\to\infty} r(\xi, \bar{D}_k) \geqq r(\xi, \hat{D})$, (3.32) cannot hold. Hence (3.29) holds and $\hat{D}$ is a Bayes solution.

For any probability measure ξ on Ω one of the following three conditions must hold:

(1) $\text{Min}_d\, W(\xi, d) < r(\xi, D)$ for any D for which $n(x, D) \geqq 1$.

(2) $\text{Min}_d\, W(\xi, d) \leqq r(\xi, D)$ for all D for which $n(x, D) \geqq 1$, and the equality sign holds for at least one D with $n(x, D) \geqq 1$.

(3) There exists a D with $n(x, D) \geqq 1$ such that $\text{Min}_d\, W(\xi, d) > r(\xi, D)$.

In view of Theorem 3.6, the conditions (1), (2) and (3) can be expressed by: (1) $\rho_0(\xi) < \rho^*(\xi)$, (2) $\rho_0(\xi) = \rho^*(\xi)$ and (3) $\rho_0(\xi) > \rho^*(\xi)$, respectively.

We shall say that a probability measure ξ on Ω is of the first type if it satisfies (1), of the second type if it satisfies (2), and of the third type if it satisfies (3). Since the a posteriori probability defined in (3.1) is also a probability measure

on Ω, any a posteriori probability measure will be one of the three types mentioned above.

We shall now prove the following characterization theorem:

THEOREM 3.7.[12] *A necessary and sufficient condition for a decision function $d = D_0(x)$ to be a Bayes solution relative to a given a priori distribution ξ_0 is that the following three relations be fulfilled for any sample point x, except perhaps on a set whose probability measure is zero when ξ_0 is the a priori distribution in Ω:*

(a) *For any $m < n(x, D_0)$, the a posteriori distribution $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$ is either of the second or of the third type,*

(b) *For $m = n(x, D_0)$, the a posteriori distribution $\xi(\omega \mid \xi_0, x_1, \cdots, x_m)$ is either of the first or the second type,*

(c) *For $m = n(x, D_0)$, we have*

$$\operatorname*{Min}_d W(\xi_{x_1,\ldots,x_m}, d) = W(\xi_{x_1,\ldots,x_m}, D_0(x))$$

where $\xi_{x_1,\ldots,x_m}$ stands for an a priori distribution that is equal to the a posteriori distribution corresponding to $\xi_0, x_1, \cdots, x_m$.

PROOF: We shall omit the proof of the sufficiency of the conditions (a), (b) and (c), since it is essentially the same as that of Theorem 3.6. To prove the necessity of these conditions, let $d = D_0(x)$ be a decision function and let M^* denote the set of all sample points x for which at least one of the relations (a), (b) and (c) is violated. First, we shall show tht M^* is a set measurable (B). Let M_1^* be the set of all x's for which (a) is violated, M_2^* the set of all x's for which (b) is violated, and M_3^* the set of all x's for which (c) is violated. Clearly. M^* is shown to be measurable (B) if we can show that $M_i^*(i = 1, 2, 3)$ is measurable (B). Let $M_{ir}^*(r = 1, 2, \cdots$, ad inf) denote the subset of M_i^* for which the first violation of the corresponding condition occurs for the sample $x_1, \cdots, x_r$. We merely have to show that M_{ir}^* is measurable (B) for all i and r. The measurability of M_{3r}^* follows from the fact that $\operatorname{Min}_d W(\xi_{x_1,\ldots,x_r}, d)$ and

$$W[\xi_{x_1,\ldots,x_r}, D_0(x)]$$

are functions of x measurable (B). To show the measurability of M_{1r}^* and M_{2r}^*, it is sufficient to show that the set of all samples $x_1, \cdots, x_r$ for which $\xi_{x_1,\ldots,x_r}$ is of type $i(i = 1, 2, 3)$ is measurable (B). But this follows from the fact that $\rho_0(\xi_{x_1,\ldots,x_r})$ and $\rho^*(\xi_{x_1,\ldots,x_r})$ are functions of $(x_1, \cdots, x_r)$ measurable (B). Hence, M^* is proved to be measurable (B).

For any x in M^* let $m(x)$ be the smallest positive integer such that at least one of the relations (a), (b) and (c) is violated for the finite sample

$$x_1, x_2, \cdots, x_{m(x)}.$$

Clearly, if x is a point in M^*, then also any sample point y is in M^* for which $y_1 = x_1, \cdots, y_{m(x)} = x_{m(x)}$. Let x^0 be any particular sample point in M^* and let $r(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ denote the conditional risk when ξ_0 is the a priori

[12] See also the proof of Lemma 1 in [4].

distribution in Ω, D_0 is the decision function adopted and the first $m(x^0)$ observations are equal to $x_1^0, \cdots, x_{m(x^0)}^0$, respectively; i.e., $r(\xi_0, D_0, x_1^0, \cdots, x_{m(x^0)}^0)$ is the conditional expected value of $W(F, D_0(x)) + n(x, D_0)$, when ξ_0 is the a priori distribution in Ω, D_0 is the decision function adopted and $x_1^0, \cdots, x_{m(x^0)}^0$ are the first $m(x^0)$ observations.

Let $D_1(x)$ be the decision function determined as follows: for any x not in M^* we put $D_1(x) = D_0(x)$. For any x in M^*, let $n(x_1, D_1)$ be equal to the smallest integer $n(x) \geqq m(x)$ for which

$$\rho_0(\xi_{x_1,\ldots,x_{n(x)}}) = \rho(\xi_{x_1,\ldots,x_{n(x)}})$$

and the value of $D_1(x)$ is determined so that condition (c) of our theorem is fulfilled. Since, for any positive integer m, the subset of M^* where $m(x) = m$ is (B) measurable, $D_1(x)$ has the proper measurability properties. Applying Theorem 3.6, we see that

$$r(\xi_0, D_1, x_1, \cdots, x_{m(x)}) = \rho(\xi_{x_1,\ldots,x_{m(x)}}) \tag{3.34}$$

for any x in M^*. On the other hand, since D_0 violates at least one of the conditions (a), (b), and (c) at every point x in M^*, we have

$$r(\xi_0, D_0, x_1, \cdots, x_{m(x)}) > \rho(\xi_{x_1,\ldots,x_{m(x)}}) \tag{3.35}$$

for every x in M^*. If the probability measure of M^* is positive when ξ_0 is the a priori probability measure, the above two relations imply that

$$r(\xi_0, D_0) > r(\xi_0, D_1).$$

Thus, D_0 is not a Bayes solution and the proof of Theorem 3.7 is complete.

We shall now prove the following continuity theorem.[13]

THEOREM 3.8. *Let* $\{\xi_i\}$ $(i = 0, 1, 2, \cdots, ad\ inf.)$ *be a sequence of probability measures on* Ω *such that*

$$\lim_{i=\infty} \frac{\xi_i(\omega)}{\xi_0(\omega)} = 1 \text{ uniformly in } \omega. \tag{3.36}$$

Then

$$\lim_{i=\infty} \rho(\xi_i) = \rho(\xi_0). \tag{3.37}$$

PROOF: It follows from (3.36) that for any $\epsilon > 0$, we have for almost all values i

$$\frac{\xi_i(\omega)}{\xi_0(\omega)} < 1 + \epsilon \text{ and } \frac{\xi_0(\omega)}{\xi_i(\omega)} < 1 + \epsilon \text{ for all } \omega. \tag{3.38}$$

Our theorem is an immediate consequence of (3.38) and Theorem 3.4.

[13] A proof of this theorem for finite Ω was given by G. W. Brown and is included in [2]. See also Lemma 3 in [4].

A stronger continuity theorem is the following:

THEOREM 3.8.1. Let $\{\xi_i\}$, $(i = 0, 1, 2, \cdots, ad\ inf.)$ *be a sequence of probabitity measures on* Ω *such that*

$$\lim_{i \to \infty} \xi_i(\omega) = \xi_0(\omega)$$

uniformly in ω. *Then* (3.37) *holds.*

PROOF: It follows from (3.11) that

$$\lim_{m \to \infty} \rho_m(\xi) = \rho(\xi)$$

uniformly in ξ. Hence it is sufficient to prove that, under the conditions of the theorem,

$$\lim_{i \to \infty} \rho_m(\xi_i) = \rho_m(\xi_0)$$

for any m. Let D^m (x) denote a decision function for which n $(x, D^m) \leqq m$ for all x. It follows that, for a fixed m, $r(F, D^m)$ is bounded, uniformly in F and D^m (Assumptions 3 and 4). From the hypothesis on $\{\xi_i\}$ it then follows that

$$\lim_{i \to \infty} r(\xi_i, D^m) = r(\xi_0, D^m)$$

uniformly in D^m. From this the desired result follows readily.

A class C of probability measures ξ on Ω will be said to be convex if for any two elements ξ_1 and ξ_2 of C and for any positive value $\lambda < 1$, the probability measure $\xi = \lambda\xi_1 + (1 - \lambda)\ \xi_2$ is an element of C.

For any element d_0 of D, let C_{i,d_0} denote the class of all probability measures ξ of type i $(i = 1, 2, 3)$ for which $W(\xi, d_0) = \underset{d}{\text{Min}}\ W(\xi, d)$. Let C_d denote the set-theoretical sum of $C_{1,d}$ and $C_{2,d}$. We shall now prove the following theorem.

THEOREM 3.9. *For any element* d, *the classes* $C_{1,d}$ *and* C_d *are convex.*[14]

Let ξ_1 and ξ_2 be two elements of $C_{1,d}$. Then for any decision function $D(x)$ which requires at least one observation we have

$$W(\xi_1, d) < r(\xi_1, D) \text{ and } W(\xi_2, d) < r(\xi_2, d). \tag{3.39}$$

Let $\xi = \lambda\xi_1 + (1 - \lambda)\ \xi_2$ where λ is a positive number <1. Clearly,

$$W(\xi, d) = \lambda W(\xi_1, d) + (1 - \lambda)\ W(\xi_2, d) \tag{3.40}$$

and

$$r(\xi, D) = \lambda r(\xi_1, D) + (1 - \lambda)\ r(\xi_2, D). \tag{3.41}$$

From (3.39), (3.40) and (3.41) we obtain

$$W(\xi, d) < r(\xi, D) \quad \text{and} \quad W(\xi, d) = \underset{d^*}{\text{Min}}\ W(\xi, d^*). \tag{3.42}$$

Hence ξ is an element of $C_{1,d}$ and the convexity of $C_{1,d}$ is proved. The convexity of C_d can be proved in the same way by replacing $<$ by $\leqq$ in (3.39) and (3.42).

[14] See also Lemma 2 in [4].

We shall say that a set L of probability measures ξ is a linear manifold if for any two elements ξ_1 and ξ_2 of L, $\xi = \alpha\xi_1 + (1 - \alpha)\,\xi_2$ is also an element of L for any real value α for which $\alpha\xi_1 + (1 - \alpha)\,\xi_2$ is a probability measure. A linear manifold L will be said to be tangent to C_d if the intersection of L and $C_{2,d}$ is not empty, but the intersection of L and $C_{1,d}$ is empty.

For any decision function $D(x)$ and for any element d of D^*, let $L(D, d)$ denote the linear manifold consisting of all ξ which satisfy the equation

$$W(\xi, d) = r(\xi, D). \tag{3.43}$$

THEOREM 3.10. *Let ξ_0 be an element of $C_{2,d}$ and let $D_0(x)$ be a decision function that requires at least one observation and is such that $W(\xi_0, d) = r(\xi_0, D_0)$. Then the linear manifold $L(D_0, d)$ is tangent to C_d.*

PROOF: ξ_0 is obviously an element of $L(D_0, d)$. Thus the intersection of $L(D_0, d)$ and $C_{2,d}$ is not empty. For any element ξ_1 of $C_{1,d}$ we have $W(\xi_1, d) < r(\xi_1, D)$ for any D that requires at least one observation. Hence, $W(\xi_1, d) < r(\xi_1, D_0)$ and, therefore, ξ_1 cannot be an element of $L(D_0, d)$. This proves our theorem.

4. Applications to the case where Ω and D^* are finite. In this section we shall apply the general results of the preceding section to the following special case: the space Ω consists of a finite number of elements, $F_1, \cdots, F_k$ (say), and the space D^* consists of the elements $d_1, \cdots, d_k$ where d_i denotes the decision to accept the hypothesis H_i that F_i is the true distribution. Let

$$W(F_i, d_j) = W_{ij} = 0 \text{ for } i = j \text{ and } >0 \text{ for } i \neq j. \tag{4.1}$$

It will be sufficient to discuss the cases $k = 2$ and $k = 3$, since the extension to $k > 3$ will be obvious. We shall first consider the case $k = 2$. In this case any a priori distribution ξ is represented by two numbers g_1 and g_2 where g_i is the a priori probability that F_i is true $(i = 1, 2)$. Thus, $g_i \geqq 0$ and $g_1 + g_2 = 1$. Let ξ_i denote the a priori distribution corresponding to $g_i = 1$ $(i = 1, 2)$. Clearly C_{d_1} contains ξ_1 but not ξ_2, and C_{d_2} contains ξ_2 but not ξ_1. Because of Theorems 3.9 and 3.7, C_{d_1} and C_{d_2} are closed and convex. Furthermore, we obviously have

$$g_2W_{21} \leqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_1} \tag{4.2}$$

and

$$g_2W_{21} \geqq g_1W_{12} \text{ for all } \xi \text{ in } C_{d_2}. \tag{4.3}$$

Let $\xi_0 = (g_1^0, g_2^0)$ be the a priori distribution for which

$$g_2^0W_{21} = g_1^0W_{12}. \tag{4.4}$$

It follows from (4.2) and (4.3) that there exist two positive numbers c' and c'' such that

$$0 < c' \leqq g_2^0 \leqq c'' < 1 \tag{4.5}$$

and such that the class C_{d_1} consists of all ξ for which $g_2 \leqq c'$, and the class C_{d_2} consists of all ξ for which $g_2 \geqq c''$.

Thus, the following decision procedure will be a Bayes solution relative to the a priori distribution $\xi = (g_1, g_2)$: *If $g_2 \leqq c'$ or $\geqq c''$, do not take any observations and make the corresponding final decision. If $c' < g_2 < c''$, continue taking observations until the a posteriori probability of H_2 is either $\geqq c''$ or $\leqq c'$. If this a posteriori probability is $\geqq c''$, accept H_2, and if it is $\leqq c'$, accept H_1.*

The a posteriori probability of H_2 after the first m observations have been made is given by

$$g_{2m} = \frac{g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{g_1 p(x_1 \mid F_1) \cdots p(x_m \mid F_1) + g_2 p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}. \tag{4.6}$$

If $c' < g_2 < c''$ and if the probability (under F_1 as well as under F_2) is zero that $g_{2m} = c'$ or $= c''$ for some m, then it follows from Theorem 3.8 that the above described Bayes solution is essentially unique; i.e., any other Bayes solution can differ from the one given above only on a set whose probability measure is zero under both F_1 and F_2.

Provided that at least one observation is made, one can easily verify that the above described Bayes solution is identical with a sequential probability ratio test for testing H_2 against H_1. The sequential probability ratio test is defined as follows (see [3]): Two positive constants A and B $(B < A)$ are chosen. Experimentation is continued as long as the probability ratio

$$\frac{p_{2m}}{p_{1m}} = \frac{p(x_1 \mid F_2) \cdots p(x_m \mid F_2)}{p(x_1 \mid F_1) \cdots p(x_m \mid F_1)} \tag{4.7}$$

satisfies the inequality $B < \frac{p_{2m}}{p_{1m}} < A$. If $\frac{p_{2m}}{p_{1m}} \geqq A$, accept H_2. If $\frac{p_{2m}}{p_{1m}} \leqq B$, accept H_1. The Bayes solution described above coincides with this probability ratio test for properly chosen values of the constants A and B.

The results described above for $k = 2$ are essentially the same as those contained in Lemmas 1 and 2 of an earlier publication [4] of the authors.

We shall now discuss the case $k = 3$. Any a priori distribution ξ can be represented by a point with the barycentric coordinates g_1, g_2 and g_3, where g_i is the a priori probability of $H_i (i = 1, 2, 3)$. The totality of all possible a priori distributions ξ will fill out the triangle T with the vertices 0_1, 0_2 and 0_3 where 0_i represents the a priori distribution corresponding to $g_i = 1$ (see Figure 1).

Clearly, the vertex 0_i is contained in C_{d_i}. Thus, because of Theorem 3.9, $C_{d_i} (i = 1, 2, 3)$ is a convex subset of T containing the vertex 0_i, as indicated in Figure 1.

If one of the components of $\xi = (g_1, g_2, g_3)$ is zero, say $g_i = 0$, then H_i can be disregarded and the problem of constructing Bayes solutions reduces to the previously considered case where $k = 2$. Thus, in particular, the determination of the boundary points $P_1, P_2, \cdots, P_6$ of C_{d_1}, C_{d_2} and C_{d_3} which are on the boundary of the triangle T, reduces to the previously considered case, $k = 2$.

It follows from Theorems 3.8 and 3.9 that the intersection of C_{d_i} with any straight line T_i through 0_i is a closed segment. One endpoint of this segment is, of course, 0_i. Let B_i denote the other endpoint. It follows from Theorem 3.7 that B_i must be a point of C_{2,d_i}. Any interior point of 0_iB_i can be shown to be an element of C_{1,d_i}. The proof of this is very similar to that of Theorem 3.9.

We shall now show how tangents to the sets C_{d_1}, C_{d_2} and C_{d_3} can be constructed at the boundary points $P_1, P_2, \cdots, P_6$. Consider, for example, the boundary point P_1 of C_{d_1} that lies on the line $0_1 0_2$. Let ξ_1 be the a priori distribution represented by the point P_1. Since the a priori probability of H_3 is zero according to ξ_1, we can disregard H_3 in constructing Bayes solutions relative to ξ_1. Let $D_1(x)$ be a sequential probability ratio test for testing H_1 against H_2

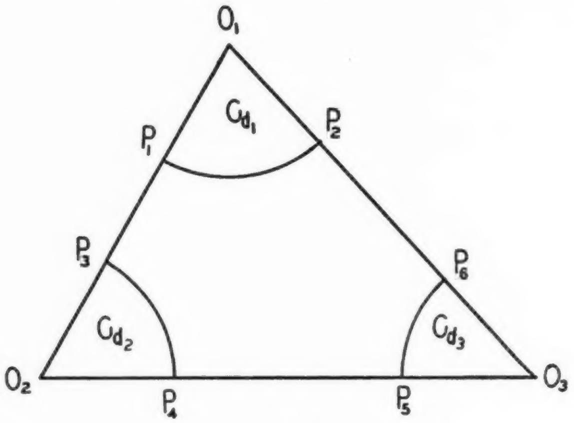

FIG. 1

which requires at least one observation and which is a Bayes solution relative to ξ_1. Since ξ_1 is a boundary point, such a decision function D_1 exists. Thus, we have

$$W(\xi_1, d_1) = r(\xi_1, D_1) = \operatorname*{Inf}_{D} r(\xi_1, D). \tag{4.8}$$

Let α_{ij} denote the probability of accepting H_j when H_i is true and D_1 is the decision function adopted. Let, furthermore, n_i denote the expected number of observations required by the decision procedure when F_i is true and D_1 is adopted. Then, for any a priori distribution $\xi = (g_1, g_2, g_3)$ we have

$$r(\xi, D_1) = \sum_{i,j} g_i W_{ij} \alpha_{ij} + \sum_i g_i n_i \tag{4.9}$$

and

$$W(\xi, d_1) = \sum_i g_i W_{i1}. \tag{4.10}$$

Thus, the linear manifold $L(D_1, d_1)$ is simply the straight line given by the equation

$$\sum_i g_i W_{i1} = \sum_{i,j} g_i W_{ij} \alpha_{ij} + \sum_i g_i n_i . \tag{4.11}$$

This straight line goes through P_1 and, because of Theorem 3.10, it is tangent to C_{d_1}. Tangents at the same points $P_2, \cdots, P_6$ can be constructed in a similar way.

The convexity properties of the sets $C_{d_i} (i = 1, 2, \cdots, k)$ were established by the authors prior to the more general results described in Sections 2 and 3 and were stated by one of the authors in an address given at the Berkeley meeting of the Institute of Mathematical Statistics, June, 1948. More general results when Ω and D^* are finite, admitting also non-linear cost functions, were obtained later by Arrow, Blackwell and Girshick [2].

REFERENCES

[1] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947), pp. 279–313.

[2] K. J. Arrow, D. Blackwell, M. A. Girshick, "Bayes and minimax solutions of sequential decision problems," *Econometrica*, Vol. 17 (1949), pp. 213–244.

[3] A. Wald, "*Sequential analysis*," John Wiley & Sons, New York, 1947.

[4] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 326–339.

[5] S. Saks, "*Theory of the integral*," Hafner Publishing Company, New York.

[6] A. Wald, "Statistical decision functions," *Annals of Math. Stat.*, Vol. 20 (1949), pp 165–205.

ON THE DISTRIBUTIONS OF MIDRANGE AND SEMI-RANGE IN SAMPLES FROM A NORMAL POPULATION

K. C. S. PILLAI

University of Travancore, Trivandrum.

1. Summary. In this paper the simultaneous distribution of midrange and semi-range has been obtained and used to derive the distributions of midrange and semi-range in samples taken from a normal population.

2. Introduction. The concept of ordering a sample has given rise to innumerable problems for statistical investigation. Several authors have contributed to the study of ordered individuals and, in particular, to the study of extreme individuals, their sum and difference in samples from a normal population. L. H. C. Tippett [1] has studied the first four moments of the range and has tabled the mean-range for sample size ranging from two to thousand. Student [2] has determined the nature of the distribution of range for particular sample sizes by purely empirical methods. T. Hojo [3] has compared the standard error of midrange to that of median and mean in normal samples. E. S. Pearson and H. O. Hartley [4] have tabled the values of the probability integral of range for sample size up to twenty. E. J. Gumbel [5], [6], [7] has established the independence of the extreme values in large samples from population of unlimited range and obtained the distributions of range and midrange. The asymptotic distribution of range has also been investigated by G. Elfving [8]. J. F. Daly [9] has devised a t-test adopting range in place of standard deviation in Student's t and in a modified t-test E. Lord [10] has used range instead of standard deviation. An extension to two populations of an analogue of Student's t-test using the sample range has been worked out by John E. Walsh [11]. S. S. Wilks [12] has given a complete and detailed account of the researches on order statistics and also a number of suggestions regarding possibilities of utilising order statistics in statistical inference. In this paper the distribution of midrange has been developed as a series and a method of evaluating the probability integral for semi-range based on an infinite series expansion for the normal probability integral has been suggested.

3. Distributions of midrange and semi-range. Let

$$x_1 \leq x_2 \cdots \leq x_n$$

be an ordered sample from a normal population with zero mean and unit standard deviation. Then the joint distribution of x_1 and x_n, the lowest and highest values respectively, is given by [13],

$$(1) \qquad p(x_1, x_n) = [n(n-1)/2\pi]\left[\int_{x_1}^{x_n} e^{-t^2/2}\, dt/\sqrt{2\pi}\right]^{n-2} e^{-(x_1^2+x_n^2)/2}.$$

Let

$$M = (x_1 + x_n)/2$$

and

$$W = (x_n - x_1)/2.$$

M is the midrange and W is the semi-range of the sample. From (1) the simultaneous distribution of M and W reduces to

$$(2) \qquad p(M, W) = [n(n-1)/\pi]e^{-(M^2+W^2)}\left[\int_{M-W}^{M+W} e^{-t^2/2}\, dt/\sqrt{2\pi}\right]^{n-2}.$$

It has been shown [14] that if

$$(3) \qquad F(M, W) = \left[\int_{M-W}^{M+W} e^{-t^2/2}\, dt/\sqrt{2\pi}\right]^{k},$$

$$(4) \qquad F(M, W) = e^{-k(M^2+W^2)/2}[A_0^{(k)} + A_1^{(k)}M^2 + \cdots + A_i^{(k)}M^{2i} + \cdots],$$

where $A_i^{(k)}$ coefficient is given by

$$(5) \qquad \begin{aligned} 2iA_i^{(k)} = kA_{i-1}^{(k)} - k\sqrt{2/\pi}[A_{i-1}^{(k-1)}W \\ + A_{i-2}^{(k-1)}W^3/\Gamma(4) + \cdots + A_0^{(k-1)}W^{2i-1}/\Gamma(2i)]. \end{aligned}$$

Using expansion (4) equation (2) reduces to

$$(6) \qquad p(M, W) = [n(n-1)/\pi]e^{-n(M^2+W^2)/2} \sum_{i=0}^{\infty} A_i^{(n-2)} M^{2i}.$$

It is evident that the A's involve terms of the form

$$[\phi(W)]^s W^q e^{-mW^2/2}$$

where s, q, m are positive integers and

$$\phi(W) = \sqrt{2/\pi} \int_0^W e^{-t^2/2}\, dt.$$

Integrating (6) with respect to W

$$(7) \qquad p(M) = [n(n-1)/\pi]e^{-nM^2/2} \sum_{i=0}^{\infty} B_i M^{2i}$$

where

$$(8) \qquad B_0 = \sqrt{\pi/2}\, I(n-2, 0, 2),$$

$$(9) \qquad \begin{aligned} B_1 &= [(n-2)/2][\sqrt{\pi/2}\, I(n-2, 0, 2) - I(n-3, 1, 3)], \\ B_2 &= [(n-2)/2^2\Gamma(3)][\sqrt{\pi/2}\, (n-2)I(n-2, 0, 2) \end{aligned}$$

$$(10) \qquad - (2n - 5)I(n - 3, 1, 3) - (1/3)I(n - 3, 3, 3) + \sqrt{2/\pi}\,(n - 3)I(n - 4, 2, 4)]$$

where

$$(11) \qquad I(s, q, m) = \sqrt{2/\pi} \int_0^\infty [\phi(x)]^s x^q e^{-mx^2/2}\, dx.$$

Using the method of integration by parts, the evaluation of $I(s, q, m)$ can be reduced ultimately to that of $I(p, 0, r)$ and this function for different values of p and r is given in Table I.

TABLE I

Values of Integrals $I(p, 0, r)$[1]

p	r = 2	4	6	8
1	0.277,063,21	0.147,583,62	0.100,735,97	0.076,490,19
2	0.152,980,4	0.064,094,20	0.037,255,93	0.025,060,53
3	0.098,373	0.033,453,6	0.016,808,71	
4	0.069,10	0.019,535,1	0.008,589,57	
5	0.051,44	0.012,325,5		
6	0.039,90	0.008,223,9		
7	0.031,94			
8	0.026,17			

The first five B Coefficients for n ranging from 3 to 10 are tabled below.

TABLE II

Values of B Coefficients.

n	B_0	B_1	B_2	B_3	B_4
3	0.347,247,25	0.040,642,87	0.002,772,90	0.000,133,80	0.000,005,00
4	0.191,732	0.058,751	0.010,906	0.001,460	0.000,153
5	0.123,292	0.067,184	0.021,526	0.004,988	0.000,909
6	0.086,60	0.070,93	0.033,23	0.011,20	0.002,97
7	0.064,47	0.072,20	0.045,65	0.020,28	0.007,14
8	0.050,01	0.072,09	0.057,22	0.032,21	0.014,59
9	0.040,03	0.071,27	0.068,95	0.047,01	0.024,98
10	0.032,80	0.069,97	0.080,31	0.064,66	0.040,51

[1] The integrals have been evaluated by using (14).

The accuracy obtained by keeping the first five terms in $p(M)$ may be judged from the following values of the total probability calculated for small values of n.

TABLE III.

Total probability keeping the first five terms in $p(M)$

Size of sample	3	4	5	6	7
Total probability	0.999,998	0.999,92	0.999,56	0.998,8	0.997,8

Integrating (6) with respect to M, $p(W)$ may be obtained. But $p(W)$ involves integral $\phi(W)$ and to evaluate the integral probability of W expansions for $\phi(W)$ and its powers have to be developed.

Since $$\phi(W) = \sqrt{2/\pi}\int_0^W e^{-t^2/2}\,dt = \sqrt{2/\pi}W(1 - W^2/6 + \cdots),$$

a convenient expansion is given by

$$\sqrt{2/\pi}\int_0^W e^{-t^2/2}\,dt = \sqrt{2/\pi}\,We^{-W^2/6}(1 + a_2W^4 + \cdots + a_iW^{2i} + \cdots) \tag{12}$$

where a_i follows the recurrence relation

$$3(2i + 1)a_i - a_{i-1} = (-1)^i/3^{i-1}\Gamma(i + 1), \tag{13}$$

as may be seen by differentiating (12) with respect to W and equating the coefficient of W^{2i} on both sides. Again

$$[\phi(W)]^j = (2/\pi)^{j/2}e^{-jW^2/6}W^jS^j \tag{14}$$

where

$$S = 1 + a_2W^4 + a_3W^6 + \cdots + a_iW^{2i} + \cdots \tag{15}$$

and

$$S^j = 1 + K_2^{(j)}W^4 + K_3^{(j)}W^6 + \cdots \tag{16}$$

where

$$K_i^{(j)} = \sum_{s=1}^{i} {}^jC_s\, s!\,a_1^{s_1}a_2^{s_2}\cdots a_i^{s_i}/s_1!\,s_2!\cdots s_i! \tag{17}$$

and

$$\begin{aligned} s_1 + 2s_2 + \cdots + is_i &= i,\\ s_1 + s_2 + \cdots + s_i &= s. \end{aligned} \tag{17a}$$

Clearly $a_i = K_i^{(1)}$. In evaluating the $K_i^{(j)}$'s summation with respect to s is first

performed, the values of $s_1, s_2, \cdots, s_i$ being obtained so as to satisfy the relations (17a); and thereafter the values of the a's are substituted. It may be noted that $a_1 = 0$. The K coefficients for j up to 8 and i up to 13 are given below.

TABLE IV

$K_i^{(j)}$ Coefficients.

j	i = 2	3	4	5
1	0.011,111,11	−0.0^{3}35,273,369	0.0^{4}44,091,711	−0.0^{5}17,814,833
2	0.022,222,22	−0.0^{3}70,546,737	0.0^{3}21,164,021	−0.0^{4}11,401,493
3	0.033,333,33	−0.0^{2}10,582,011	0.0^{3}50,264,550	−0.0^{4}28,860,029
4	0.044,444,44	−0.0^{2}14,109,348	0.0^{3}91,710,758	−0.0^{4}54,157,091
5	0.055,555,56	−0.0^{2}17,636,684	0.0^{2}14,550,265	−0.0^{4}87,292,680
6	0.066,666,67	−0.0^{2}21,164,021	0.0^{2}21,164,021	−0.0^{3}12,826,680
7	0.077,777,78	−0.0^{2}24,691,358	0.0^{2}29,012,346	−0.0^{3}17,707,944
8	0.088,888,89	−0.0^{2}28,218,695	0.0^{2}38,095,238	−0.0^{3}23,373,061

j	i = 6	7	8	9
1	0.0^{6}10,087,459	−0.0^{8}38,065,882	0.0^{9}14,772,299	−0.0^{11}47,770,889
2	0.0^{5}13,059,860	−0.0^{7}78,306,957	0.0^{8}57,379,607	−0.0^{9} 32,240,604
3	0.0^{5}49,870,764	−0.0^{6}35,414,321	0.0^{7}37,246,865	−0.0^{8} 26,934,251
4	0.0^{4}12,515,888	−0.0^{6}96,195,746	0.0^{6}13,039,809	−0.0^{7} 10,793,811
5	0.0^{4}25,264,163	−0.0^{5}20,323,918	0.0^{6}33,614,797	−0.0^{7} 30,234,979
6	0.0^{4}44,603,642	−0.0^{5}36,960,883	0.0^{6}72,070,037	−0.0^{7} 68,563,784
7	0.0^{4}71,905,926	−0.0^{5}60,836,892	0.0^{5}13,654,992	−0.0^{6} 13,526,252
8	0.0^{3}10,854,319	−0.0^{5}93,258,365	0.0^{5}23,672,301	−0.0^{6} 24,174,891

j	i = 10	11	12	13
1	0.0^{12}14,640,444	−0.0^{14}40,268,872	0.0^{15}10,359,029	−0.0^{17}24,535,539
2	0.0^{10}18,330,114	−0.0^{12}91,351,579	0.0^{13}43,595,840	−0.0^{14}19,132,452
3	0.0^{9} 21,506,514	−0.0^{10}14,469,203	0.0^{12}96,661,910	−0.0^{13}58,727,628
4	0.0^{8} 10,849,591	−0.0^{10}87,178,260	0.0^{11}72,767,557	−0.0^{12}54,213,617
5	0.0^{8} 36,260,639	−0.0^{9} 32,719,538	0.0^{10}32,219,900	−0.0^{11}27,049,719
6	0.0^{8} 95,092,297	−0.0^{9} 93,120,388	0.0^{9} 10,472,881	−0.0^{11}96,020,717
7	0.0^{7} 21,247,442	−0.0^{8} 22,112,968	0.0^{9} 27,825,332	−0.0^{10}27,369,553
8	0.0^{7} 42,365,199	−0.0^{8} 46,218,579	0.0^{9} 64,147,144	−0.0^{10}66,862,484

Using (12) the probability integral for W can be evaluated with the help of tables of Incomplete Gamma Functions.

REFERENCES

[1] L. H. C. Tippet, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925),pp. 364–387.

[2] Student, "Errors in routine analysis," *Biometrika*, Vol. 19 (1927), pp. 151–164.

[3] T. Hojo, "Distribution of the median, quartiles and interquartile distance in samples from a normal population," *Biometrika*, Vol. 23 (1931), pp. 315–360.

[4] E. S. Pearson and H. O. Hartley, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), pp. pp. 301–310.

[5] E. J. Gumbel, "Ranges and midranges," *The Annals of Math. Stat.*, Vol. 15 (1944), No. 4, pp. 414–422.

[6] E. J. Gumbel, "On the independence of the extremes in a sample," *The Annals of Math. Stat.*, Vol. 17 (1946), No. 1, pp. 78–80.

[7] E. J. Gumbel, "The distribution of the range," *The Annals of Math. Stat.*, Vol. 18 (1947), No. 3, pp. 384–412.

[8] G. Elfving, "The asymptotical distribution of range in samples from a normal population," *Biometrika*, Vol. 34, (1947), pp. 111–119.

[9] J. F. Daly, "On the use of the sample range in an analogue of Student's t-test," *The Annals of Math. Stat.*, Vol. 17 (1946), No. 1, pp. 71–74.

[10] E. Lord, "The use of range in place of standard deviation in the t-test," *Biometrika*, Vol. 34 (1947), pp. 41–67.

[11] J. E. Walsh, "An extension to two populations of an analogue of Student's t-test using the sample range," *Annals of Math. Stat.*, Vol. 18 (1947), No. 2, pp. 280–285.

[12] S. S. Wilks, "Order Statistics," *Am. Math. Soc. Bull.*, Vol. 54 (1948), pp. 6–50.

[13] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, Princeton, 1943, p. 90.

[14] K. C. S. Pillai, "A note on ordered samples," *Sankhya*, Vol. 8 (1948), Part 4, pp. 375–380.

THE IMPOSSIBILITY OF CERTAIN SYMMETRICAL BALANCED INCOMPLETE BLOCK DESIGNS

By S. S. Shrikhande

University of North Carolina

Introduction and Summary. An arrangement of v varieties or treatments in b blocks of size k, $(k < v)$, is known as a balanced incomplete block design if every variety occurs in r blocks and any two varieties occur together in λ blocks. These parameters obviously satisfy the equations

$$bk = vr \tag{1}$$

$$\lambda(v - 1) = r(k - 1). \tag{2}$$

Fisher [1] has also proved that the inequality

$$b \geq v, \qquad r \geq k \tag{3}$$

must hold. If v, b, r, k and λ are positive integers satisfying (1), (2) and (3), then a balanced incomplete block design with these parameters possibly exists, but the actual existence of a combinatorial solution is not ensured. These conditions are thus necessary but not sufficient for the existence of a design. Fisher and Yates in their tables [2] have listed all designs with $r \leq 10$ and given combinatorial solutions, where known. A balanced incomplete block design in which $b = v$, and hence $r = k$ is called a symmetrical balanced incomplete block design. The impossibility of the symmetrical designs with parameters $v = b = 22$, $r = k = 7$, $\lambda = 2$ and $v = b = 29$, $r = k = 8$, $\lambda = 2$ was first demonstrated by Hussain [3], [4] essentially by the method of enumeration. The object of the present note is to give an alternative simple proof of the impossibility of these designs and to show that the only unknown remaining symmetrical design in Fisher and Yates' tables, *viz.* $v = b = 46$, $r = k = 10$, $\lambda = 2$, is definitely impossible. Symmetrical designs with $\lambda \leq 5$, $r, k \leq 20$, which are impossible combinatorially, are also listed.

1. A necessary condition for the existence of a symmetrical balanced incomplete block design when v is even.

Theorem 1. *A necessary condition for the existence of a symmetrical balanced incomplete block design with parameters v, r and λ, where v is even, is that $r - \lambda$ be a perfect square.*

Proof. Let $N = (n_{ij})$ be a square matrix of v rows and v columns where

$$n_{ij} = 1 \text{ or } 0 \tag{4}$$

according as the i-th treatment does or does not occur in the j-th block. Put

$$B = NN' \tag{5}$$

Since every treatment occurs in r blocks and every pair of treatments in λ blocks, we have, if the design is possible,

$$B = \begin{pmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \cdots & & \cdots & \cdot \\ \lambda & \lambda & \cdots & r \end{pmatrix}. \tag{6}$$

Subtracting the first column from all the other columns and then adding to the first row all the other rows, we see that

$$\begin{aligned} |B| &= [r + \lambda(v-1)](r-\lambda)^{v-1} \\ &= r^2(r-\lambda)^{v-1} \text{ from (2)}. \end{aligned} \tag{7}$$

But from (5)

$$|B| = |N|^2.$$

Since $|N|$ is integral, it follows that $(r - \lambda)^{v-1}$ is the square of an integer, and hence if v is even, $r - \lambda$ must be a perfect square.

COROLLARY. *The following symmetrical designs are impossible.*

(A_1)	$v = b = 22$	$r = k = 7$	$\lambda = 2$
(A_2)	$v = b = 46$	$r = k = 10$	$\lambda = 2$
(A_3)	$v = b = 92$	$r = k = 14$	$\lambda = 2$
(A_4)	$v = b = 106$	$r = k = 15$	$\lambda = 2$
(A_5)	$v = b = 172$	$r = k = 19$	$\lambda = 2$
(A_6)	$v = b = 34$	$r = k = 12$	$\lambda = 4.$

As already mentioned in the introduction, the impossibility of (A_1) has been proved by Hussain [3], but for the design (A_2) it was hitherto unknown whether or not a solution is possible and it was left as a blank in the latest edition of Fisher and Yates' tables.

2. Application of method of Bruck and Ryser.

In a recent paper Bruck and Ryser [5] have proved the impossibility of some finite projective planes with the help of the properties of matrices whose elements are integers. Their method is immediately applicable to our own problem.

Let A and B be two symmetric matrices of order n with elements in the rational field. The matrices A and B are congruent, written $A \sim B$, provided there exists a nonsingular matric C with elements in the rational field, such that $A = C'BC$. The congruence of matrices satisfies the usual requirements of an "equals" relationship.

If A is an integral symmetric matrix of order n and rank n, we can always construct an integral diagonal matrix $D = (d_1, \cdots, d_n)$, where $d_i \neq 0$, $i = 1, 2, \cdots n$ such that $D \sim A$. The number of negative terms i, called the index of A, is an invariant by Sylvester's Law.

Define $d = (-1)^i\delta$ where δ is the square-free positive part of $|A|$. Then since $|B| = |C|^2|A|$, d is another invariant of A.

Now let A be a nonsingular and symmetric integral matrix of order n. Let D_r be the leading principal minor determinant of order r and suppose that $D_r \neq 0$ for $r = 1, 2, \cdots n$. Define

$$C_p(A) = (-1, -D_n)_p \prod_{j=1}^{n-1} (D_j, -D_{j+1})_p \tag{9}$$

for every odd prime p where $(m, m')_p$ is the Hilbert norm-residue symbol for arbitrary non-zero integers m and m' and for every prime p. The following two theorems are given in the collected works of Hilbert [6].

THEOREM (A). *If m and m' are integers not divisible by the odd prime p, then*

$$(m, m')_p = +1 \tag{10}$$

$$(m, p)_p = (p, m)_p = (m/p), \tag{11}$$

where (m/p) is the Legendre symbol. Moreover, if $m \equiv m' \not\equiv 0 \bmod p$, then

$$(m, p)_p = (m', p)_p. \tag{12}$$

THEOREM (B). *For arbitrary non-zero integers m, m', n, n' and for every prime p,*

$$(-m, m)_p = +1 \tag{13}$$

$$(m, n)_p = (n, m)_p \tag{14}$$

$$(mm', n)_p = (m, n)_p(m', n)_p \tag{15}$$

$$(m, nn')_p = (m, n)_p(m, n')_p. \tag{16}$$

From the above it is easy to prove that for p an odd prime and every positive integer m,

$$(m, m+1)_p = (-1, m+1)_p \tag{17}$$

$$\prod_{j=1}^{m} (j, j+1)_p = ((m+1)!, -1)_p. \tag{18}$$

We can now state the fundamental Minkowski-Hasse Theorem [7].

THEOREM (C). *Let A and B be two integral symmetric matrices of order n and rank n. Suppose further that the leading principal minor determinants of A and B are different from zero. Then $A \sim B$ if and only if A and B have the same invariants i, d and C_p for every odd prime p.*

3. A necessary condition for the existence of a symmetrical balanced incomplete block design for any integer v.

Suppose the symmetrical design with parameters v, r and λ exists. Then with the previous definition of N and B,

$$B = NN' = \begin{pmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \cdots & & \cdots & \\ \lambda & \lambda & \cdots & r \end{pmatrix}.$$

Subtracting the last row from the remaining rows and then subtracting the last column from all the other columns, we get

$$Q = \begin{bmatrix} 2(r-\lambda) & (r-\lambda) & \cdots & (r-\lambda) & -(r-\lambda) \\ (r-\lambda) & 2(r-\lambda) & \cdots & (r-\lambda) & -(r-\lambda) \\ \cdots & & \cdots & & \cdots \\ (r-\lambda) & (r-\lambda) & \cdots & 2(r-\lambda) & -(r-\lambda) \\ -(r-\lambda) & -(r-\lambda) & \cdots & -(r-\lambda) & r \end{bmatrix}. \tag{19}$$

Obviously $Q \sim B$. But $B \sim I$. Hence $Q \sim I$ and, therefore, since Q and I satisfy all the conditions of Theorem C, they must have the same invariants i, d and C_p.

Let D_j denote the leading principal minor determinant of Q of order j. Then

$$D_j = (r-\lambda)^j(j+1) \qquad \text{for } j = 1, 2, \cdots, v-1 \tag{20}$$

(21) and $$D_v = |B| = r^2(r-\lambda)^{v-1}.$$

Then, omitting p for convenience,

$$C_p(Q) = (-1, -D_v)(D_{v-1}, -D_v) \prod_{j=1}^{v-2} (D_j, -D_{j+1}).$$

We use (10) $\cdots$, (18) in deriving the value of $C_p(Q)$.

Now

$$\begin{aligned}(-1, -D_v)(D_{v-1}, -D_v) &= (-1, -r^2(r-\lambda)^{v-1})((r-\lambda)^{v-1}v, -r^2(r-\lambda)^{v-1}) \\ &= (-1, -1)(-1, r^2)(-1, (r-\lambda)^{v-1})((r-\lambda)^{v-1}, r^2) \\ &\quad ((r-\lambda)^{v-1}, -(r-\lambda)^{v-1})(v, r^2)(v, -(r-\lambda)^{v-1}) \\ &= (-1, (r-\lambda)^{v-1})(v, -(r-\lambda)^{v-1}) \\ &= (-1, (r-\lambda)^{v-1})(v, -1)(v, (r-\lambda))^{v-1}.\end{aligned}$$

Also

$$\begin{aligned}\prod_{j=1}^{v-2} (D_j, -D_{j+1}) &= \prod_{1}^{v-2} ((r-\lambda)^j(j+1), -(r-\lambda)^{j+1}(j+2)) \\ &= \left\{\prod_{1}^{v-2} ((r-\lambda)^j, -(r-\lambda)^{j+1})(j+1, -(j+2))\right\} S \\ &= S \prod_{1}^{v-2} ((r-\lambda)^j, -(r-\lambda)^j)((r-\lambda)^j, (r-\lambda))(j+1, j+2)(j+1, -1) \\ &= S \prod_{1}^{v-2} ((r-\lambda), (r-\lambda))^j(j+2, -1)(j+1, -1) \\ &= S \prod_{1}^{v-2} (r-\lambda, -1)^j(j+2, -1)(j+1, -1) \\ &= S(r-\lambda, -1)^{(v-1)(v-2)/2}((v-1)!, -1)(v!, -1) \\ &= S(r-\lambda, -1)^{(v-1)(v-2)/2}(v, -1),\end{aligned}$$

where

$$S = \prod_{1}^{v-2} ((r-\lambda)^j, j+2)((r-\lambda)^{j+1}, j+1)$$
$$= \prod_{1}^{v-2} ((r-\lambda)^j, j+2)((r-\lambda)^{j-1}, j+1)$$
$$= \prod_{1}^{v-2} ((r-\lambda)^j, j+2) \prod_{j=0}^{v-3}((r-\lambda)^j, j+2)$$
$$= (r-\lambda, v)^{v-2}.$$
$$\therefore C_p(Q) = (r-\lambda, -1)^{v(v-1)/2} (v, -1)^2 (r-\lambda, v)^{2v-3}$$
$$= (r-\lambda, -1)^{v(v-1)/2}(r-\lambda, v)^{2v-3}. \tag{22}$$

Hence we can enunciate the following theorem:

THEOREM 2. *A necessary condition for the existence of a symmetrical balanced incomplete block design with parameters v, r and λ is that*

$$C_p(Q) = (r-\lambda, -1)_p^{v(v-1)/2}(r-\lambda, v)_p^{2v-3} = +1$$

for all odd prime p, where $(m, n)_p$ is the Hilbert norm-residue symbol.

When v is even we have seen that a necessary condition for the existence of the design is that $r - \lambda$ be a perfect square. Then it is easily seen that

$$C_p(Q) = +1$$

for all odd prime p. Therefore, even if the design is really non-existent, its impossibility cannot be proved by this method.

When, however, v is odd we can in many instances demonstrate the impossibility of the design.

Consider the design

$$(A_7) \qquad v = b = 29, \qquad r = k = 8, \qquad \lambda = 2.$$
$$C_p(Q) = (6, -1)_p^{29.14}(6, 29)_p^{55}$$
$$= (3, 29)_p(2, 29)_p$$
$$= (29/3) \text{ for p} = 3$$
$$= (2/3) \text{ for p} = 3$$
$$= -1 \quad \text{for p} = 3.$$

Hence the design (A_7) is impossible. As mentioned in the introduction, the impossibility has already been demonstrated by Hussain [4] by a rather lengthy method amounting to a complete exhaustion of all possibilities. The following designs with $\lambda \leq 5$ and r, $k \leq 20$ can be similarly proved to be impossible by applying Theorem 2.

(A_8)	$v = b = 137$	$r = k = 17$	$\lambda = 2$
(A_9)	$v = b = 67$	$r = k = 12$	$\lambda = 2$
(A_{10})	$v = b = 103$	$r = k = 18$	$\lambda = 3$
(A_{11})	$v = b = 53$	$r = k = 13$	$\lambda = 3$
(A_{12})	$v = b = 43$	$r = k = 15$	$\lambda = 5$
(A_{13})	$v = b = 77$	$r = k = 20$	$\lambda = 5$.

My thanks are due to Professor R. C. Bose under whose guidance this research was carried out.

REFERENCES

[1] R. A. Fisher, "An examination of the different possible solutions of a problem in incomplete blocks," *Annals of Eugenics*, Vol. 10 (1940), pp. 52–75.

[2] R. A. Fisher and F. Yates, *Statistical Tables for Biological Agricultural and Medical Research* (1949), Hafner Publishing Company, New York.

[3] Q. M. Hussain, "Impossibility of the symmetrical incomplete block design with $\lambda = 2$, $k = 7$," *Sankhyā*, Vol. 7 (1946), pp. 317–322.

[4] Q. M. Hussain, "Symmetrical incomplete block design with $\lambda = 2$, $k = 8$ or 9," *Bulletin of the Calcutta Mathematical Society*, XXXVII (1945), pp. 115–123.

[5] R. H. Bruck and H. J. Ryser, "Non-existence of certain finite projective planes," *Canadian Jour. Math.*, Vol. 1 (1949), pp. 88–93.

[6] D. Hilbert, *Gesammelte Abhandlungen* I (Berlin, 1932), pp. 161–173.

[7] H. Hasse, "Über die Äquivalenz quadratischer formen im körper rationalen zahlen," *J. Reine Angew. Math.*, Vol. 152 (1923), pp. 205–224.

NOTES

This section is devoted to brief research and expository articles and other short items.

THE SAMPLING DISTRIBUTION OF THE RATIO OF TWO RANGES FROM INDEPENDENT SAMPLES[1]

Richard F. Link

University of Oregon

Let us consider a sample of n ordered observations $(x_1 < x_2 < \cdots < x_n)$ drawn from a population with variance σ^2. Let $w = (x_n - x_1)/\sigma$. Let us consider the joint sampling distribution of w_1 and w_2 for two samples, not necessarily the same size, drawn from populations with the same variance. If the two samples were drawn independently, then the joint sampling distributions of w_1 and w_2 may be written as the product of the sampling distributions of w_1 and w_2.

If we make the change of variable $r = w_1/w_2$, $w_2 = w$, and if w is integrated over its range of definition, the cumulative distribution of the ratio of two ranges remains. This may be written as

$$F(R) = \int_0^R dr \int_0^\infty dw \cdot w \cdot h_2(w) \cdot h_1(wr), \tag{1}$$

where h_1 is the pdf for w_1 and h_2 is the pdf for w_2.

To obtain more explicit results, specific distribution functions may be considered. The following table gives the sampling distribution of the ratio of two ranges from independent samples for the indicated density functions $f(x)$. Notice that for the normal distribution it was possible to obtain results only for some special cases.

In Table 1 for $F(R)$, w_1 and w_2 represent ranges computed from samples of size n_1 and n_2 respectively.

Notice that formula (1) for $F(R)$ is equivalent to the following expressions

$$Pr(w_1/w_2 < R) = F(R) = \int_0^\infty dw_2 \int_0^{Rw_2} dw_1\, h(w_1) \cdot h(w_2).$$

The region of integration for the last expression is simply the region in the w_2, w_1 plane to the right of the line $w_1 = Rw_2$.

This integration was done numerically. Table 2 gives values of R for all combinations of n_1 and $n_2 \leq 10$ and for $\alpha = .005, .01, .025, .05, .10$ such that

$$Pr(w_1/w_2 < R) = \alpha$$

where w_1 and w_2 are ranges computed from samples of size n_1 and n_2 drawn from

[1] This work was done under contract N6onr-218/IV with the Office of Naval Research.

normal populations with the same variance. It is believed that these values are correct to within one place in the last reported figure.

These tabled values may be used as critical values for testing the hypothesis that two independent samples were drawn from normal populations with the same variance. This test is therefore comparable to the F test. Some sort of

TABLE 1

$f(x)$	$F(R) = Pr\,(w_1/w_2 < R)$
$1 \quad 0 \le x \le 1$ $0 \quad$ all other x	$n_2(n_2-1)\,R^{n_1-1}\left[\dfrac{n_1}{n_1+n_2-2} - \dfrac{(n_1+R[n_1-1])}{n_1+n_2-1} + \dfrac{R(n_1-1)}{n_1+n_2}\right]$
$e^{-x} \quad 0 \le x < \infty$ $0 \quad x < 0$	$1-(n_1-1)(n_2-1)\sum_{i=0}^{n_1-2}\sum_{j=0}^{n_2-2}\binom{n_1-2}{i}\binom{n_2-2}{j} \cdot \dfrac{(-1)^{i+j}}{[1+j+(1+i)R]\,(1+i)}$
$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \quad -\infty < x < \infty$	$n_1 = 2, \quad n_2 = 2 \quad \dfrac{2}{\pi}\tan^{-1} R$ $n_1 = 2, \quad n_2 = 3 \quad \dfrac{6}{\pi}\tan^{-1}\dfrac{R}{\sqrt{4+3R^2}}$ $n_1 = 3, \quad n_2 = 2 \quad \dfrac{6}{\pi}\left(\tan^{-1}\sqrt{3+4R^2}-\dfrac{\pi}{3}\right)$ $n_1 = 3, \quad n_2 = 3$ $\int_0^R dr\left[\dfrac{27r}{2\pi^2}\left\{\dfrac{2}{r^2}(u\tan^{-1}u - v\tan^{-1}v) + \dfrac{1}{6r^2(1+r^2)}(w\tan^{-1}w - u\tan^{-1}u) + \dfrac{u^2y}{r}\tan^{-1}2ry\right\}\right]$ where $u = [3(r^2+1)]^{-\frac{1}{2}} \qquad w = (7r^2+3)^{-\frac{1}{2}}$ $v = (4r^2+3)^{-\frac{1}{2}} \qquad y = (3r^2+4)^{-\frac{1}{2}}$

measure of the relative performance of these two tests seems desirable. An attempt to measure the performance of this test relative to the F test was made by comparing the tolerance intervals of the distribution of this ratio with those of the F test.

The length of the interval containing the central $1 - 2\alpha$ proportion of the distribution of F was compared with a similar length for the distribution of w_1/w_2 for $n_1 = n_2 = n$. The square of the ratio of these lengths will be called δ_α^2.

TABLE 2

$$Pr\left(\frac{w_1}{w_2} < R\right) = .005$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0078	.0052	.0043	.0039	.0038	.0037	.0036	.0035	.0034
3	.096	.071	.059	.054	.051	.048	.045	.042	.041
4	.21	.16	.14	.13	.12	.12	.11	.11	.10
5	.30	.24	.22	.20	.19	.18	.18	.17	.16
6	.38	.32	.28	.26	.25	.24	.23	.22	.22
7	.44	.38	.34	.32	.30	.29	.28	.27	.26
8	.49	.43	.39	.36	.35	.33	.32	.31	.30
9	.54	.47	.43	.40	.38	.37	.36	.35	.34
10	.57	.50	.46	.44	.42	.40	.39	.38	.37

$$Pr\left(\frac{w_1}{w_2} < R\right) = .01$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.0157	.0105	.0080	.0070	.0068	.0066	.0063	.0062	.0061
3	.136	.100	.084	.079	.073	.069	.065	.062	.060
4	.26	.20	.18	.17	.16	.15	.14	.14	.13
5	.38	.30	.26	.24	.23	.22	.21	.21	.20
6	.46	.37	.33	.31	.29	.28	.27	.26	.26
7	.53	.43	.39	.36	.34	.33	.32	.31	.30
8	.59	.49	.44	.41	.39	.37	.36	.35	.34
9	.64	.53	.48	.45	.43	.41	.40	.39	.38
10	.68	.57	.52	.49	.46	.45	.43	.42	.41

$$Pr\left(\frac{w_1}{w_2} < R\right) = .025$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.039	.026	.019	.018	.017	.016	.016	.015	.015
3	.217	.160	.137	.124	.115	.107	.102	.098	.095
4	.37	.28	.25	.23	.21	.20	.19	.18	.18
5	.50	.39	.34	.32	.30	.28	.27	.26	.25
6	.60	.47	.42	.38	.36	.34	.33	.32	.31
7	.68	.54	.48	.44	.42	.40	.38	.37	.36
8	.74	.59	.53	.49	.46	.44	.43	.42	.41
9	.79	.64	.57	.53	.50	.48	.47	.46	.44
10	.83	.68	.61	.57	.54	.52	.50	.49	.48

TABLE 2—*Continued*

$$Pr\left(\frac{w_1}{w_2} < R\right) = .05$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.079	.052	.039	.036	.034	.032	.031	.030	.028
3	.31	.23	.20	.18	.16	.15	.14	.14	.13
4	.50	.37	.32	.29	.27	.26	.25	.24	.23
5	.62	.49	.42	.40	.36	.35	.33	.32	.31
6	.74	.57	.50	.46	.43	.41	.40	.38	.37
7	.80	.64	.57	.52	.49	.47	.45	.44	.43
8	.86	.70	.62	.57	.54	.51	.50	.48	.47
9	.91	.75	.67	.61	.58	.55	.53	.52	.51
10	.95	.80	.70	.65	.61	.59	.57	.55	.54

$$Pr\left(\frac{w_1}{w_2} < R\right) = .10$$

n_1 \ $n_2 =$	2	3	4	5	6	7	8	9	10
2	.158	.105	.077	.074	.069	.066	.062	.059	.056
3	.46	.33	.28	.25	.23	.22	.21	.20	.19
4	.67	.49	.42	.38	.36	.34	.32	.31	.30
5	.84	.62	.53	.48	.45	.43	.41	.39	.38
6	.97	.72	.62	.56	.52	.50	.48	.46	.45
7	1.07	.80	.69	.63	.59	.56	.54	.52	.50
8	1.15	.87	.75	.68	.64	.61	.58	.56	.54
9	1.21	.92	.80	.73	.68	.65	.62	.60	.58
10	1.26	.98	.85	.77	.72	.68	.66	.64	.62

TABLE 3

n	*Relative precision of the range as an estimate of σ*	δ^2
2	1.00	1.00
3	.99	.99
4	.98	.97
5	.96	.95
6	.93	.92
7	.91	.90
8	.89	.89
9	.87	.88
10	.85	.86

For statistics having normal sampling distributions such a ratio would be independent of α and would be equivalent to the ratio of the variances of these sampling distributions. It was found that δ_α^2 is independent of α except for a maximum change of 1 in the second decimal for the values of $\alpha = .005, .01, .025, .05, .10$. These values of δ^2 are presented in Table 3 along with the relative precision of the range as an estimate of σ as given by Mosteller [1].

It is interesting to note that δ^2 corresponds very closely to the relative precision of the range as an estimate of σ.

REFERENCE

[1] F. Mosteller, "On some useful 'inefficient' statistics," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 377–408.

A NOTE ON THE ESTIMATION OF A DISTRIBUTION FUNCTION BY CONFIDENCE LIMITS

By Frank J. Massey, Jr.

University of Oregon

Let $F(x)$ be the continuous cumulative distribution function of a random variable X, and let $x_1 < x_2 < x_3 < \cdots < x_n$ be the results of n independent observations on X arranged in order of size. We wish to estimate $F(x)$ by means of the band $S_n(x) \pm \lambda/\sqrt{n}$ where $S_n(x)$ is defined by

$$S_n(x) = \begin{cases} 0 & \text{if } x < x_1, \\ k/n & \text{if } x_k \leq x < x_{k+1}, \\ 1 & \text{if } x \geq x_n. \end{cases}$$

Thus we wish to know the probability, say $P_n(\lambda)$, that the band is such that $S_n(x) - \dfrac{\lambda}{\sqrt{n}} < F(x) < S_n(x) + \dfrac{\lambda}{\sqrt{n}}$ for all x. This problem has been previously studied [1] [2] [3] [4] [5] and a limiting distribution has been obtained [1] [4] [5] and tabled [3] [4]. However apparently no error terms for the limiting distribution, or practical methods of obtaining $P_n(\lambda)$ have been given. Such a method is given here.

It has been shown [2] that $P_n(\lambda)$ is independent of $F(x)$ provided only that $F(x)$ is continuous, and thus it is sufficient to consider only the case

$$F(x) = \begin{cases} 0 \text{ if } x < 0, \\ x \text{ if } 0 \leq x \leq 1, \\ 1 \text{ if } x \geq 1. \end{cases}$$

We will find the probability that $S_n(x)$ falls wholly in the band $F(x) \pm k/n$ (here $\lambda = k/\sqrt{n}$) where k is an integer or a rational number, and intermediate values may be obtained by interpolation. To illustrate the method we shall assume that k is an integer.

Divide the interval (0, 1) into n parts by the points $1/n, 2/n, \cdots, (n - 1)/n$. The step function $S_n(x)$ rises by jumps of exactly $1/n$. Thus, in order to be inside the band at $x = i/n$, $S_n(x)$ would have to pass through exactly one of the lattice points whose ordinates are $(i - k + 1)/n, (i - k + 2)/n, \cdots, (i + k - 1)/n$.

Suppose that the step function stays inside the band by means of α_i of the observations falling in the interval $\left(\frac{i-1}{n}, \frac{i}{n}\right)$ $i = 1, 2, \cdots, n$. The a priori probability of this happening is given by the multinomial law as

$$P_r(\alpha_1 \cdots \alpha_n) = \frac{n!}{\alpha_1!\,\alpha_2! \cdots \alpha_n!}\left(\frac{1}{n}\right)^{\alpha_1}\left(\frac{1}{n}\right)^{\alpha_2}\cdots\left(\frac{1}{n}\right)^{\alpha_n}$$
$$= \frac{1}{\alpha_1! \cdots \alpha_n!}\frac{n!}{n^n}$$

since $\sum_1^n \alpha_i = n$.

Thus the probability of the step function staying in the band is given by

$$P_n(\lambda) = \sum \frac{n!}{n^n}\frac{1}{\alpha_1!\,\alpha_2! \cdots \alpha_n!} = \frac{n!}{n^n}\sum\frac{1}{\alpha_1! \cdots \alpha_n!}$$

where the summation is over all possible combinations of $\alpha_1, \cdots, \alpha_n$ such that $\max_x | S_n(x) - x | < \frac{\lambda}{\sqrt{n}}$ and $\sum_{i=1}^{n} \alpha_i = n$.

Let $U_i(m) = \sum_i \frac{1}{\alpha_1! \cdots \alpha_m!}$, $i = 1, 2, \cdots, 2k - 1$ be the sum of all the terms indicated such that $S_n(x)$ arrives at the lattice point $\left(\frac{m}{n}, \frac{m-k+i}{n}\right)$ by a route that stays inside the band. Since the $S_n(x)$ is non-decreasing it can only pass through a point

$$\left(\frac{m+1}{n}, \frac{m-k+1+j}{n}\right), \qquad m = 0, 1, \cdots, n-1; \qquad j = 1, 2, \cdots, 2k-1,$$

if it previously passed through one of the points

$$\left(\frac{m}{n}, \frac{m-k+1}{n}\right)\cdots, \left(\frac{m}{n}, \frac{m-k+2}{n}\right)\cdots, \left(\frac{m}{n}, \frac{m-k+j+1}{n}\right).$$

If it passed through $\left(\frac{m}{n}, \frac{m-k+h}{n}\right)$ the value of α_{m+1} would have to be $(j + 1 - h)$ and the product $U_h(m)\frac{1}{(j+1-h)!}$ would be part of $U_j(m + 1)$. This is true for all $h = 1, 2, \cdots, j + 1$ and all of these terms would give different paths for $S_n(x)$ so we have

$$U_j(m+1) = \sum_{h=1}^{j+1} \frac{1}{(j+1-h)!}U_h(m), \qquad j = 1, 2, \cdots, 2k-1,$$

where it is understood $U_h(m) = 0$ if $h \geq m + k$.

Thus we have a set of $2k - 1$ linear homogeneous difference equations. They may be reduced to a single difference equation by eliminating $2k - 2$ of the variables by substitution. This results in the following difference equation.

$$\sum_{h=1}^{2k-1} (-1)^h \frac{(2k-h)^h}{h!} U_k(2k - 1 - h + m) = 0.$$

TABLE 1

k	$n = 5$	10	20	25	30	35	40	45
1.0	.0384	.0004						
1.5	.3276	.0449						
2.0	.6521	.2513	.0238					
2.5	.8880	.5139						
3.0	.9699	.7331	.2955					
3.5	.9947	.8522						
4.0	.99935	.9410	.6473					
5.0		.9922	.8624	.7637	.6629	.5674	.4808	.4042
6.0		.9994	.9569	.9057	.8420	.7725	.7016	.6322
7.0			.9892	.9683	.9359	.8945	.8471	.7962
8.0			.9979	.9911	.9774	.9566	.9295	.8974
9.0			.9997	.9979	.9931	.9842	.9708	.9529

k	$n = 50$	55	60	65	70	75	80
5.0	.3377	.2807	.2324	.1918	.1577	.1294	.1060
6.0	.5662	.5046	.4478	.3954	.3492	.3072	.2696
7.0	.7439	.6916	.6403	.5908	.5435	.4987	.4566
8.0	.8616	.8234	.7837	.7434	.7031	.6633	.6244
9.0	.9312	.9063	.8789	.8496	.8189	.7874	.7554

Initial conditions on either the simultaneous equations or on the single equation are

$$U_i(0) = 0 \text{ for } i \neq k,$$

$$U_k(0) = 1 \text{ for } i = k.$$

After values of $U_k(n)$ have been found the value of $P_n\left(\frac{k}{\sqrt{n}}\right)$ can be found by multiplying $U_k(n)$ by $\frac{n!}{n^n}$.

The values of $U_k(n)$ can be obtained numerically either from the simultaneous

equations or from the single equation. Table 1 was computed partly by numerical solution of the simultaneous equations above and partly by setting up similar equations connecting $U_i(x + 5)$ to $U_t(x)$, $t = 1, 2, \cdots, i + 5$. Either method could be set up on punch cards if an extensive table was desired. Notice that to get $U_k(n)$ all $U_k(t)$, $t = 1, 2, \cdots, n - 1$ are also found. Table 1 gives some computed values of $P_n(k)$. Table 2 gives results interpolated from Table 1, showing the approach of $P_n(\lambda)$ to its limiting distribution.

If the width of the band is $2\left(\frac{k}{l}\right)$ when k and l are integers a similar procedure to that above can be used. However instead of dividing the interval (0, 1) into n parts it is necessary to divide it into $l \cdot n$ parts.

TABLE 2

n	$\lambda = .9$	1.0	1.10	1.20	1.30	1.40
10	.66	.78	.85	.91	.95	.97
20	.65	.77	.85	.91	.94	.97
30	.65	.76	.85	.90	.94	.96
40	.64	.76	.84	.90	.94	.96
50	.64	.75	.84			
60	.63	.75	.84			
70	.63	.75	.83			
80	.63	.74				
∞	.607	.730	.822	.888	.932	.960

It has been suggested (2) that instead of a band bounded by $y = x \pm c$ it might be convenient to use a band bounded by the lines $y = px + q$ and $y = p'x + q'$. If $p = p'$ and if p, q, q' are rational the probability of $S_n(x)$ staying inside the band can be evaluated by the method presented above. If $p \neq p'$ and if p, p', q, q' are all rational a similar procedure could be used but it would be very tedious.

[1] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Recueil Math. de Moscou*, Vol. 6 (1939), pp. 3–26.

[2] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[3] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathematique de l'Universite de Moscou*, Vol. 2 (1939), fasc. 2.

[4] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Ist. Ital. att.* Giorn. Vol. 4 (1933), p. 1–11.

[5] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distribution" *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–190.

SIGNIFICANCE LEVELS FOR A k-SAMPLE SLIPPAGE TEST[1]

FREDERICK MOSTELLER AND JOHN W. TUKEY

Harvard University and Princeton University

1. Summary. Mosteller has recently [1, 1948] proposed a k-sample slippage test and has given percentage points for selected n, k and r for the case of k equal samples of size n. When the samples are of unequal size, exact significance levels can be calculated very quickly from

$$P_r = \frac{\sum n_i^{(r)}}{N^{(r)}} \text{ where } x^{(r)} = x(x-1)\cdots(x-r+1),$$

by the method explained in section 3 below.

The significance values for k equal samples of $n \geq 10$ are very well approximated by

$$P_r = \frac{1}{k^{r-1}} e^{-r(r-1)(k-1)/2N}$$

where $N = kn$.

A convenient rough approximation for unequal samples may be given in terms of k^*, an "effective" number of samples, which is given by

$$k^* = \frac{(\sum_i n_i)^2}{\sum n_i^2},$$

the one-sided significance level will then be approximately given by

$$P_r = (k^*)^{-(r-1)}.$$

This approximation can be easily applied with the aid of Table 1. Thus, for example, with four samples of sizes 7, 5, 5, 2, we have

$$k^* = \frac{(7+5+5+2)^2}{49+25+25+4} = \frac{361}{103} = 3.50,$$

whence from the table $r = 3$ lies at a one-sided level approximately between 5% and 10%, $r = 4$ approximately between 1% and 2.5%, $r = 5$ between 0.5% and 1%, $r = 6$ near 0.2%, and so on. Direct calculation yields 5.7%, 1.2%, 0.2% and 0.03%. The approximation is, in this example, quite satisfactory for moderate significance levels and conservative for more extreme significance levels.

2. Derivation. The statistic considered by Mosteller is the number of cases in one sample greater than all cases in all the $k - 1$ other samples. We derive its distribution briefly.

Since the statistic depends only on the order of the $n_1 + n_2 + \cdots + n_k = N$

[1] Prepared in connection with research sponsored by the Office of Naval Research.

values, we can consider the actual values taken on to be fixed, and consider their allotment to the various samples. Assuming all of them to come from a single continuous distribution, we may consider these fixed values to be all distinct, and any way of allotting them to labelled places in the various samples as equally likely.

Consider the r largest values. They can all be allotted to places in the i-th sample in $n_i^{(r)} = n_i(n_i - 1) \cdots (n_i - r + 1)$ ways, and to arbitrary places in $N^{(r)}$ ways. Thus they will be allotted to some single sample in the fraction

$$P_r = \frac{\sum_i n_i^{(r)}}{N^{(r)}}$$

of all cases. This is clearly the probability that Mosteller's statistic is r or more.

TABLE 1

Approximate critical values of k^* for various levels of significance

One-sided level	10%	5%	2.5%	1%	0.5%	0.2%	0.1%
Two-sided level	20%	10%	5%	2%	1%	0.4%	0.2%
$r = 2$	10.0	20.0	40.0	100.0	200.0	500.0	1000.0
$r = 3$	3.2	4.5	6.3	10.0	14.1	22.4	31.6
$r = 4$		2.7	3.4	4.6	5.8	7.9	23.0
$r = 5$			2.5	3.2	3.8	4.7	5.6
$r = 6$				2.5	2.9	3.5	4.0
$r = 7$						2.8	3.2
$r = 8$							2.6

3. Unequal samples—an exact computation. Our practical problem is to compute P_r for small values of r and a fixed set of n_i. If we recognize the numerators as the unnormalized factorial moments of the distribution of sample sizes, we see that the computation goes smoothly according to the scheme shown in Table 2 (where the columns of multipliers $n - 1$, $n - 2$, $n - 3$, etc. may be partially covered for convenience during the computation.): For example: $132 = 11(12)$, $1320 = 10(132)$, $\cdots$ $42 = 6(7)$. The numbers in the last line of Table 2 give successively the percentages $100\, P_1$, $100\, P_2$, $\cdots$. Of course $P_1 = 1$ because some sample must have the largest value. It is clear that exact computation for any reasonable set of n_i is quite easy.

4. Equal samples—an approximation. In the case of k equal samples, we have

$$P_r = \frac{kn^{(r)}}{N^{(r)}}.$$

Let us try to approximate to $n^{(r)}$ by expansion in powers. We have

$n^{(r)} = n(n-1)\cdots(n-r+1) = n^r(1-1/n)(1-2/n)\cdots(1-(r-1)/n)$, so that

$$\begin{aligned}\log n^{(r)} &= r\log n + \sum_{x=1}^{r-1}\log(1-x/n)\\ &= r\log n - \sum_{x=1}^{r-1}(x/n + x^2/2n^2 + x^3/3n^3\cdots)\\ &= r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2 + 0(n^{-3}),\end{aligned}$$

TABLE 2

Sample Computation

for $\{n_i\} = (12, 11, 11, 11, 10, 10, 10, 10, 9, 9, 7, 4)$

$n-2$	$n-1$	n	f	nf	$n^{(2)}f$	$n^{(3)}f$
10	11	12	1	12	132	1320
9	10	11	3	33	330	2970
8	9	10	4	40	360	2880
7	8	9	2	18	144	1008
5	6	7	1	7	42	210
2	3	4	1	4	12	24
$N-2$	$N-1$	N	Sums	114	1020	8412
112	113	114	$N^{(r)}$	114	12882	1442784
			P_r	100%	7.9%	0.58%

and hence

$$\begin{aligned}\log P_r &= \log k + \log n^{(r)} - \log N^{(r)} = \log k + \log n^{(r)} - \log(nk)^{(r)}\\ &= \log k + r\log n - r(r-1)/2n - r(r-1)(2r-1)/12n^2\\ &\qquad - r\log nk + r(r-1)/2nk + r(r-1)(2r-1)/12n^2k^2 + 0(n^{-3})\\ &= -(r-1)\log k - \frac{r(r-1)}{2n}\left(1 - \frac{1}{k} + \frac{2r-1}{6n} - \frac{2r-1}{6nk^2} + 0(n^{-2})\right).\end{aligned}$$

We get the following three approximations:

(1) $$P_r \doteq \frac{1}{k^{r-1}};$$

and noting that $\dfrac{1-1/k}{n} = \dfrac{k-1}{kn} = \dfrac{k-1}{N}$,

(2) $$P_r \doteq k^{-(r-1)}\, e^{(-r(r-1)(k-1))/2N} = \frac{1}{[ke^{(r(k-1))/2N}]^{r-1}};$$

and finally

(3) $$P_r \doteq k^{-(r-1)} e^{(-r(r-1)(k-1)/2N)(1+(2r-1)/6n)}.$$

5. Comparison of results. The results obtained with various equal sample approximations will be compared with the exact values for several cases. The effective number of samples, k^*, used with (1), (2), and (3), is computed from

$$k^* = \frac{(\sum n_i)^2}{\sum n_i^2},$$

a formula which is often an easy and effective way to allow for different sizes of samples.

TABLE 3

Comparison of Approximations

Sizes of Samples	N	k	r	P_r in exact	(1)	(2)	(3)	(4)
10, 10, 10, 10	40	4.00	2	23.08	25.00	23.19	23.13	≤25.00
			3	4.85	6.25	4.99	4.80	≤6.25
7, 5, 5, 2	19	3.50+	2	24.56	28.53	25.01	24.82	≤28.53
			3	5.67	8.14	5.48	5.18	≤8.76
12, 11, 11, 11								
10, 10, 10, 10	114	11.46	2	7.92	8.73	7.96	7.96	≤8.73
9, 9, 7, 4			3	0.58	0.76	0.58	0.56	≤0.78

A fourth approximation, which always gives a conservative estimate of the significance of the result is obtained by replacing $n^{(r)}$ by n^r throughout, this gives

(4) $$P_r = \frac{\sum n_i^r}{N^r},$$

which is equivalent to approximation (1) when the samples are of equal size, or when $r = 2$.

The results are shown in Table 3.

Thus it seems clear that either (1) or (4) are good enough for rough work. The choice will depend on which formula one prefers to remember. The amount of work is about the same for either method. When something better is required the exact method of section 3 seems appropriate. Indeed some may prefer it to any approximation.

REFERENCE

[1] FREDERICK MOSTELLER, "A k-sample slippage test for an extreme population," *Annals of Math. Stat.*, Vol. 19 (1948), p. 58–65.

ADJUSTMENT OF AN INVERSE MATRIX CORRESPONDING TO A CHANGE IN ONE ELEMENT OF A GIVEN MATRIX

By Jack Sherman and Winifred J. Morrison

The Texas Company Research Laboratories, Beacon, New York

1. Introduction. Many methods have been published in recent years for carrying out the numerical computation of the inverse of a matrix [1], [2]. In all these methods, the amount of computation increases rapidly with increase in order of the matrix.

The utility of a computational method for obtaining the inverse of a matrix would be increased considerably if the inverse could be transformed in a simple manner, corresponding to some specified change in the original matrix, thus eliminating the necessity of computing the new inverse from the beginning. The problem that is considered in the present paper is one of changing one element in the original matrix, and of computing the resulting changes in the elements of the new inverse directly from those of the old inverse.

2. Computational method. Let

a_{ij}, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, n$ denote the elements of an nth order square matrix $\boldsymbol{a}$;

b_{ij}, denote the elements of $\boldsymbol{b}$, the inverse of $\boldsymbol{a}$;

A_{ij}, denote the elements of $\boldsymbol{A}$ which differs from $\boldsymbol{a}$ only in one element, say A_{RS};

B_{ij}, denote the elements of $\boldsymbol{B}$, the inverse $\boldsymbol{A}$.

Let

$$A_{RS} = a_{RS} + \Delta a_{RS}.$$

The set of equations by means of which $\boldsymbol{B}$ may be computed from Δa_{RS} and $\boldsymbol{b}$ is

$$B_{rj} = b_{rj} - \frac{b_{rR}\, b_{Sj}\,\Delta\, a_{RS}}{1 + b_{SR}\,\Delta\, a_{RS}}, \qquad \begin{array}{l} r = 1, 2, \cdots, n, \\ j = 1, 2, \cdots, n, \end{array} \tag{1}$$

provided that $1 + b_{SR}\Delta a_{RS} \neq 0$.

The validity of equation (1) may be demonstrated by multiplying through by A_{ir}, $(r = 1, 2, \cdots, n)$ and adding the results:

$$\sum_{r=1}^{n} A_{ir}\, B_{rj} = \sum_{r=1}^{n} A_{ir}\, b_{rj} - \frac{b_{Sj}\,\Delta\, a_{RS}}{1 + b_{SR}\,\Delta\, a_{RS}} \sum_{r=1}^{n} A_{ir}\, b_{rR}, \tag{2}$$
$$(i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, n).$$

Consider separately the equations for which $i \neq R$, and for which $i = R$.

Case I. $i \neq R$. By hypothesis, $A_{ir} = a_{ir}$ for $i \neq R$. Hence equations (2) become

$$\sum_{r=1}^{n} A_{ir}\, B_{rj} = \sum_{r=1}^{n} a_{ir}\, b_{rj} - \frac{b_{Sj}\,\Delta\, a_{RS}}{1 + b_{SR}\,\Delta\, a_{RS}} \sum_{r=1}^{n} a_{ir}\, b_{rR}, \tag{3}$$
$$(i = 1, 2, \cdots, R-1, R+1, \cdots, n;\ j = 1, 2, \cdots, n).$$

The last sum vanishes because $\boldsymbol{a}$ and $\boldsymbol{b}$ are inverse matrices, and hence

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} a_{ir} b_{rj} \qquad (i = 1, 2, \cdots, R-1, R+1, \cdots, n; j = 1, 2, \cdots, n). \tag{4}$$

Case II. $i = R$. Equation (2) becomes

$$\sum_{r=1}^{n} A_{Rr} B_{rj} = \sum_{r=1}^{n} A_{Rr} b_{rj} - \frac{b_{Sj} \Delta a_{RS}}{1 + b_{SR} \Delta a_{RS}} \sum_{r=1}^{n} A_{Rr} b_{rR} \quad (j = 1, 2, \cdots, n). \tag{5}$$

In each of the summations, there will be a term for which $r = S$, in which case $A_{RS} = a_{RS} + \Delta a_{RS}$. In all other cases, $A_{Rr} = a_{Rr}$. Hence (5) can be written as

$$\begin{aligned} \sum_{r=1}^{n} A_{Rr} B_{rj} &= \sum_{r=1}^{n} a_{Rr} b_{rj} + \Delta a_{RS} b_{Sj} \\ &- \left(\frac{b_{Sj} \Delta a_{RS}}{1 + b_{SR} \Delta a_{RS}}\right)\left(\sum_{r=1}^{n} a_{Rr} b_{rR} + \Delta a_{RS} b_{SR}\right) \qquad (j = 1, 2, \cdots, n). \end{aligned} \tag{6}$$

Since $\boldsymbol{a}$ and $\boldsymbol{b}$ are inverse matrices, the second summation on the right-hand side of (6) is equal to unity, and hence (6) becomes

$$\sum_{r=1}^{n} A_{Rr} B_{rj} = \sum_{r=1}^{n} a_{Rr} b_{rj} \qquad (j = 1, 2, \cdots, n). \tag{7}$$

The sets of equations (4) and (7) can be written as one set of equations:

$$\sum_{r=1}^{n} A_{ir} B_{rj} = \sum_{r=1}^{n} a_{ir} b_{rj} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, n), \tag{8}$$

and hence $\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$.

3. Illustrative numerical example. In actual applications, equations (1) are conveniently subdivided into three groups, namely, those for which $r = S$, those for which $j = R$, and all others. In the first two cases, these reduce to

$$B_{Sj} = \frac{b_{Sj}}{1 + b_{SR} \Delta a_{RS}}, \qquad (j = 1, 2, \cdots, n), \tag{9}$$

$$B_{rR} = \frac{b_{rR}}{1 + b_{SR} \Delta a_{RS}}, \qquad (r = 1, 2, \cdots, n). \tag{10}$$

By utilizing (10), (1) becomes

$$\begin{aligned} &B_{rj} = b_{rj} - B_{rR} b_{Sj} \Delta a_{RS}, \\ &(r = 1, 2, \cdots, S-1, S+1, \cdots, n; \\ &j = 1, 2, \cdots, R-1, R+1, \cdots, n). \end{aligned} \tag{11}$$

Equations (10) and (11) show that the elements of $\boldsymbol{B}$ contained in the Sth row and Rth column are directly proportional to the corresponding elements of $\boldsymbol{b}$.

Consider

$$a = \begin{pmatrix} 2.384 & 1.238 & 0.861 & 2.413 \\ 0.648 & 1.113 & 0.761 & 0.137 \\ 1.119 & 0.643 & 3.172 & 1.139 \\ 0.745 & 2.137 & 1.268 & 0.542 \end{pmatrix}.$$

The inverse of b turns out to be

$$\mathbf{b} = \begin{pmatrix} 0.2220 & 2.5275 & -0.1012 & -1.4145 \\ -0.04806 & -0.2918 & -0.1999 & 0.7079 \\ -0.1692 & 0.01195 & 0.3656 & -0.01824 \\ 0.2801 & -2.3517 & 0.07209 & 1.0409 \end{pmatrix}.$$

Assume that a_{24} is increased by 0.4, so that

$$\mathbf{A} = \begin{pmatrix} 2.384 & 1.238 & 0.861 & 2.413 \\ 0.648 & 1.113 & 0.761 & 0.537 \\ 1.119 & 0.643 & 3.172 & 1.139 \\ 0.745 & 2.137 & 1.268 & 0.542 \end{pmatrix}.$$

Then (9), (10), and (11) become

$$B_{4j} = \frac{b_{4j}}{1 - 2.3517 \times 0.4} = 16.857\, b_{4j} \qquad (j = 1, 2, \cdots, n),$$

$$B_{r2} = 16.857\, b_{r2}\ (r = 1, 2, \cdots, n),$$

$$B_{rj} = b_{rj} - 0.4\, B_{r2} b_{4j}\ (r = 1, 2, \cdots, S-1, S+1, \cdots, n;$$
$$j = 1, 2, \cdots, R-1, R+1, \cdots, n).$$

Utilization of these equations gives

$$\mathbf{B} = \begin{pmatrix} -4.5518 & 42.608 & -1.3298 & -19.155 \\ 0.5031 & -4.9191 & -0.05805 & 2.7560 \\ -0.1919 & 0.2014 & 0.3598 & -0.1021 \\ 4.7218 & -39.644 & 1.2153 & 17.547 \end{pmatrix}.$$

4. Concluding remarks. It is seen from equation (1) that if $\Delta a_{RS} = -1/b_{SR}$, that is, if a_{RS} is increased by the negative of the reciprocal of the corresponding element in the transposed reciprocal matrix, then the denominator in the second term on the right-hand side of equation (1) becomes equal to zero, and $\boldsymbol{B}$ cannot be found by the present method. It is left to the reader to verify that under these conditions $\boldsymbol{A}$ is in fact singular.

In the illustrative numerical example, the denominator is only $1 - 2.3517 \times 0.4 = 0.05932$, which accounts for the large magnitude of some of the elements of $\boldsymbol{B}$. If Δa_{24} were taken to be $1/2.3517 = 0.4252$ instead of 0.4, $\boldsymbol{A}$ would have become singular.

If two or more elements in the matrix a are to be changed, the new inverse can be found by successive applications of the method.

REFERENCES

[1] H. HOTELLING, "Some new methods in matrix calculation," *Annals of Math Stat.*, Vol. 14 (1943), pp. 1-35.

[2] P. S. DWYER, "The solution of simultaneous equations," *Psychometrika*, Vol. 6 (1941), p. 101.

A CLASS OF RANDOM VARIABLES WITH DISCRETE DISTRIBUTIONS

ALBERT NOACK

Cologne, Germany

1. General results. A large class of random variables with discrete probability distributions can be derived from certain power series. Let

$$f(z) = \sum_{x=0}^{\infty} a_x z^x, \quad a_x \text{ real}, \quad |z| < r.$$

We may have either non-negative coefficients a_x or we may have $(-1)^x a_x \geq 0$. In the first case take $0 < z < r$; and in the second case take $-r < z < 0$. Define a random variable with the distribution

$$P\{\xi = x\} = \frac{a_x z^x}{f(z)}; \qquad x = 0, 1, 2, \cdots. \tag{1}$$

The above conditions insure $P\{\xi = x\} \geq 0$ for all x; besides

$$\sum_x P\{\xi = x\} = \frac{1}{f(z)} \sum_x a_x z^x = 1.$$

The distribution of ξ may be called the power series distribution (p.s.d.). The mean of such a distribution is

$$E(\xi) = \sum_x x P\{\xi = x\} = \frac{1}{f(z)} \sum_x x a_x z^x.$$

Hence it follows that

$$E(\xi) = z \frac{f'(z)}{f(z)} = z \frac{d}{dz} \log f(z). \tag{2}$$

We have for the moments about the origin

$$\mu_r' = \sum_x x^r P\{\xi = x\} = \frac{1}{f(z)} \sum_x x^r a_x z^x,$$

and hence

$$z \frac{d\mu_r'}{dz} = \frac{1}{f(z)} \sum_x x^{r+1} a_x z^x - z \frac{f'(z)}{f(z)} \frac{1}{f(z)} \sum_x x^r a_x z^x.$$

Thus we have the recurrence relation

$$\mu'_{r+1} = z \frac{d\mu'_r}{dz} + \mu'_1 \mu'_r . \tag{3}$$

The central moments are

$$\mu_r = \sum_x (x - \mu'_1)^r P\{\xi = x\} = \frac{1}{f(z)} \sum_x (x - \mu'_1)^r a_x z^x,$$

and hence

$$z \frac{d\mu_r}{dz} = \frac{1}{f(z)} \sum_x x(x - \mu'_1)^r a_x z^x - zr \frac{d\mu'_1}{dz} \frac{1}{f(z)} \sum_x (x - \mu'_1)^{r-1} a_x z^x$$

$$- z \frac{f'(z)}{f(z)} \cdot \frac{1}{f(z)} \sum_x (x - \mu'_1)^r a_x z^x.$$

The sum of the first and third term will be found to be μ_{r+1}, hence

$$z \frac{d\mu_r}{dz} = \mu_{r+1} - rz \frac{d\mu'_1}{dz} \mu_{r-1} ,$$

whence we have for the central moments of a p.s.d. the recurrence relation

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + r \frac{d\mu'_1}{dz} \mu_{r-1} \right]. \tag{4}$$

Putting $r = 1$, $\mu_0 = 1$, $\mu_r = 0$, we get the variance of ξ

$$\mu_2 = \sigma^2(\xi) = z \frac{d\mu'_1}{dz} = z^2 \frac{d^2}{dz^2} \log f(z) + \mu'_1 = z^2 \frac{f''(z)}{f(z)} - z^2 \left[\frac{f'(z)}{f(z)}\right]^2 + z \frac{f'(z)}{f(z)}. \tag{5}$$

By (5), (4) assumes the form

$$\mu_{r+1} = z \frac{d\mu_r}{dz} + r\mu_2 \mu_{r-1} . \tag{4'}$$

The characteristic function of ξ is

$$\varphi(t) = \sum_x e^{itx} P\{\xi = x\} = \frac{1}{f(z)} \sum_x a_x e^{itx} z^x,$$

or

$$\varphi(t) = \frac{f(e^{it} z)}{f(z)}. \tag{6}$$

To get a relation connecting the cumulants κ_n and the moments μ'_r about the origin, we differentiate both sides of the identity

$$\sum_{r=1}^{\infty} \frac{\kappa_r}{r!} (it)^r = \log \sum_{\rho=0}^{\infty} \frac{\mu'_\rho}{\rho!} (it)^\rho$$

with respect to (it), identifying coefficients in $(it)^{r-1}$ we get[1]

$$\mu_r' = \sum_{j=1}^{r} \binom{r-1}{j-1} \mu_{r-j}' \kappa_j . \tag{7}$$

Differentiation of (7) with respect to z gives

$$\frac{d\mu_r'}{dz} = \sum_{j=1}^{r} \binom{r-1}{j-1} \left[\frac{d\mu_{r-j}'}{dz} \kappa_j + \mu_{r-j}' \frac{d\kappa_j}{dz} \right]. \tag{7'}$$

Substitution of (7) and (7′) in (3) gives

$$\sum_{j=1}^{r+1} \binom{r}{j-1} \mu_{r+1-j}' \kappa_j = \sum_{j=1}^{r} \binom{r-1}{j-1} \left\{ \left[z \frac{d\mu_{r-j}'}{dz} + \mu_1' \mu_{r-j}' \right] \kappa_j + z\mu_{r-j}' \frac{d\kappa_j}{dz} \right\},$$

or by (3) after a little re-arrangement

$$\kappa_{r+1} = z \sum_{j=1}^{r} \binom{r-1}{j-1} \mu_{r-j}' \frac{d\kappa_j}{dz} - \sum_{j=2}^{r} \binom{r-1}{j-2} \mu_{r+1-j}' \kappa_j . \tag{8}$$

2. Special cases.

(a) Choosing $f(z) = e^z$, ξ has Poisson-distribution

$$P\{\xi = x\} = \frac{z^x e^{-z}}{x!}. \tag{1a}$$

(2) and (5) are the well known relations $E(\xi) = \sigma^2(\xi) = z$; the recurrence formula (4) assumes the form[2]

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + r\mu_{r-1} \right]. \tag{4a}$$

(b) Taking $f(z) = (1 - z)^{-k}$, $k > 0$, $0 < z < 1$ we get the so-called negative binomial distribution

$$P\{\xi = x\} = \frac{\Gamma(k + x)}{x!\,\Gamma(k)} z^x (1 - z)^k, \qquad x = 0, 1, 2, \cdots . \tag{1b}$$

The mean is

$$E(\xi) = \frac{kz}{1 - z}, \tag{2b}$$

while the recurrence formula for the central moments is

$$\mu_{r+1} = z \left[\frac{d\mu_r}{dz} + \frac{rk}{(1 - z)^2} \mu_{r-1} \right], \tag{4b}$$

hence the first three moments of this distribution are

$$\sigma^2(\xi) = \mu_2 = \frac{kz}{(1 - z)^2},$$

[1] Cf. M. G. Kendall, *The Advanced Theory of Statistics*, Vol. I, p. 87.
[2] Cf. Craig, *Am. Math. Soc. Bull.*, Vol. 40 (1934), p. 262.

$$\mu_3 = \frac{kz(1+z)}{(1-z)^3}, \tag{5b}$$

$$\mu_4 = \frac{kz(1+4z+z^2+3kz)}{(1-z)^4}.$$

The characteristic function of the distribution is

$$\varphi(t) = \left(\frac{1-e^{it}z}{1-z}\right)^{-k}. \tag{6b}$$

Writing $z = \eta/(1+\eta)$, $k = h/\eta$, $\eta > 0$, $h > 0$ we get the so-called Polya-Eggenberger distribution for rare contagious events[3].

$$w\{\xi = x\} = \frac{\Gamma\left(\frac{h}{\eta}+x\right)}{x!\,\Gamma(h\eta^{-1})}\left(\frac{\eta}{1+\eta}\right)^x (1+\eta)^{-h/\eta}, \qquad x = 0, 1, 2, \cdots. \tag{1b$_1$}$$

The first four moments of this distribution are

$$\mu_1' = h \tag{2b$_1$}$$

$$\mu_2 = h(1+\eta) \tag{5b$_1$}$$

$$\mu_3 = h(1+\eta)(1+2\eta)$$

$$\mu_4 = h(1+\eta)[1+3(1+\eta)(h+2\eta)].$$

To obtain a recurrence relation for the moments consider

$$\frac{d\mu_r}{dz} = \frac{\partial\mu_r}{\partial\eta}\frac{d\eta}{dz} + \frac{\partial\mu_r}{\partial h}\frac{dh}{\partial z} = (1+\eta)^2\left[\frac{\partial\mu_r}{\partial\eta} + \frac{h}{\eta}\frac{\partial\mu_r}{\partial h}\right];$$

hence we find for this distribution by (4) and (4b)

$$\mu_{r+1} = (1+\eta)\left[\eta\frac{\partial\mu_r}{\partial\eta} + h\frac{\partial\mu_r}{\partial h} + rh\mu_{r-1}\right]. \tag{4b$_1$}$$

It follows from (4b$_1$), that μ_r is a polynomial in η and h. The characteristic function of this distribution is

$$\varphi(t) = [1+\eta(1-e^{it})]^{-h/\eta}. \tag{6b$_1$}$$

(c) The coefficients of the series $-\log(1-z) = \sum_{x=1}^{\infty} z^x/x$ are positive; the associated distribution derived is

$$P\{\xi = x\} = -\frac{z^x}{x\log(1-z)}, \qquad 0 < z < 1; \quad x = 1, 2, \cdots, \tag{1c}$$

and has the mean

$$E(\xi) = -\frac{z}{(1-z)\log(1-z)}. \tag{2c}$$

[3] Cf. *Zeits. f. angew. Math. und Mech.*, Vol. 3 (1923), p. 279–289.

Recurrence formula (4) has for this distribution the form

$$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} - r\,\frac{z + \log(1-z)}{(1-z)^2[\log(1-z)]^2}\,\mu_{r-1}\right], \tag{4c}$$

while the variance and the characteristic function of this distribution are

$$\mu_2 = \sigma^2(\xi) = -\frac{z^2 + z\log(1-z)}{(1-z)^2[\log(1-z)]^2}, \tag{5c}$$

$$\varphi(t) = \frac{\log(1-e^{it}z)}{\log(1-z)}. \tag{6c}$$

(d) The coefficients of the series $\log(1+z)/(1-z) = 2\sum_{x=1}^{\infty}(z^{2x+1})/(2x+1)$ are positive, so we can derive a random variable ξ with the distribution

$$P\{\xi = 2x+1\} = \frac{2z^{2x+1}}{(2x+1)\log\dfrac{1+z}{1-z}}, \quad 0 < z < 1, x = 1, 2, 3, \cdots. \tag{1d}$$

ξ has the mean

$$E(\xi) = \frac{2z}{(1-z^2)\log\dfrac{1+z}{1-z}}, \tag{2d}$$

the recurrence formula (4) assumes the form

$$\mu_{r+1} = z\left(\frac{d\mu_r}{dz} + 2r\cdot\frac{(1+z^2)\log\dfrac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\dfrac{1+z}{1-z}\right]^2}\,\mu_{r-1}\right), \tag{4d}$$

while the variance and the characteristic function of ξ are

$$\sigma^2(\xi) = 2z\,\frac{(1+z^2)\log\dfrac{1+z}{1-z} - 2z}{(1-z^2)^2\left[\log\dfrac{1+z}{1-z}\right]^2}, \tag{5d}$$

$$\varphi(t) = \frac{\log(1+e^{it}z) - \log(1-e^{it}z)}{\log(1+z) - \log(1-z)}. \tag{6d}$$

(e) Likewise the coefficients of the series

$$\sin^{-1} z = z + \sum_{x=1}^{\infty}\frac{1\cdot3\cdot5\cdots(2x-1)}{2\cdot4\cdot6\cdots(2x)}\,\frac{z^{2x+1}}{2x+1}$$

are positive, the derived variable ξ with the distribution

$$P\{\xi = 1\} = (\sin^{-1} z)^{-1},$$

(1e) $$P\{\xi = 2x + 1\} = \frac{1\cdot 3\cdots(2x-1)}{2\cdot 4\cdot 6\cdots(2x)}\cdot\frac{z^{2x+1}}{2x+1}(\sin^{-1} z)^{-1},$$
$$0 < z < 1, x = 1, 2, 3, \cdots,$$

has the mean

(2e) $$E(\xi) = \frac{z}{\sqrt{1-z^2}\sin^{-1} z}.$$

The recurrence formula for the moments

(4e) $$\mu_{r+1} = z\left[\frac{d\mu_r}{dz} + r\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^3(\sin^{-1} z)^2}\mu_{r-1}\right]$$

gives the variance

(5e) $$\sigma^2(\xi) = z\frac{\sin^{-1} z - z\sqrt{1-z^2}}{\sqrt{1-z^2}^3(\sin^{-1} z)^2}.$$

The characteristic function assumes the form

(6e) $$\varphi(t) = \frac{\sin^{-1} e^{it} z}{\sin^{-1} z}.$$

(f) It is well known, that series (b), (c), (d), and (e) are special cases of the hypergeometric function $F(a, b, c; z)$. This function gives a p.s.d., if $abc > 0$. If $a > 0$, $b > 0$, $c > 0$ or if $a < 0$, $b < 0$, $c > 0$, a, b integers, there exist no further restrictions on these parameters. Suppose $a < 0$, $b < 0$, $c > 0$, a integer, b not, we must have $[b] \leq a$[4]; if neither a nor b are integers, we must have $[a] = [b]$. Suppose $a < 0$, $b > 0$, $c < 0$. If c is an integer, a must be an integer $> c$. If a is an integer, but c not, we must have $[c] \leq a$. Finally if neither a nor c are integers, we must have $[a] = [c]$. Corresponding conditions are valid, if $a > 0$, $b < 0$, $c < 0$. Regarding

$$\frac{d}{dz}F(a, b; c; z) = \frac{ab}{c}F(a+1, b+1; c+1; z),$$

the mean of a random variable ξ with hypergeometric distribution is

(2f) $$E(\xi) = z\frac{ab}{c}\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}.$$

Considering the differential equation

$$z(1-z)f''(z) + [c - (a+b+1)z]f'(z) - abf(z) = 0,$$

(5) gives the variance of ξ

(5f) $$\sigma^2(\xi) = \frac{ab}{c}\cdot\frac{z}{1-z}\left\{c + [1 - c + (a+b)z]\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)} - z(1-z)\frac{ab}{c}\left[\frac{F(a+1, b+1; c+1; z)}{F(a, b; c; z)}\right]^2\right\}.$$

The higher moments of this distribution can now derived from (4′).

[4] $[b]$ means as usual the greatest integer $\leq b$.

THE GEOMETRIC RANGE FOR DISTRIBUTIONS OF CAUCHY'S TYPE

By E. J. Gumbel and R. D. Keeney

New York City and Metropolitan Life Insurance Company

1. Introduction. We consider large samples drawn from a symmetrical unlimited population whose distribution is of the Cauchy type, defined by the properties

$$\lim_{x\to\infty} x^k[1 - F(x)] = A, \qquad \lim_{x\to-\infty} (-x)^k F(x) = A, \tag{1}$$

where k and A are positive and $F(x)$ stands for the probability function. This type of distribution has no moments of an order equal to or greater than k. We construct the distribution of a certain function of the extreme values, and require only the knowledge of the type of the initial distribution, not of the distribution itself.

From each sample we pick out the largest and smallest observations, x_n and x_1. If the median of the initial distribution is zero, and the sample size is large enough, the probability of any extreme x_n or $-x_1$ being negative can be neglected. If we draw N such samples, each of large size n, we obtain N pairs of extremes, $x_{n,\nu}$ and $x_{1,\nu}$ ($\nu = 1, 2, 3, \cdots, N$). For each sample we can then compute the geometric mean, ρ, of these extremes:

$$\rho = \sqrt{x_n(-x_1)}, \tag{2}$$

which we henceforth call the *geometric* range.

The distribution of these geometric ranges can be obtained directly from the joint asymptotic distribution of the extremes. However, it is easier to obtain this distribution indirectly from the distribution of the reciprocal of the geometric range. This distribution of the reciprocal is of interest in itself: since it possesses all moments we can use it to estimate the parameters by the method of moments, whereas this problem seems to be very intricate if we start from the distribution of the geometric range itself.

2. The distribution of the reciprocal of the geometric range. The distribution of the reciprocal of the geometric range follows from a theorem of Elfving [1] which may be stated thus:

"Let x be a symmetrical unlimited variate with probability $F(x)$. Let ξ be defined by

$$\xi = 2n\sqrt{F(x_1)[1 - F(x_n)]}. \tag{3}$$

Then the asymptotic density function $g(\xi)$ and the asymptotic probability $G(\xi)$ of ξ are:

$$g(\xi) = \xi K_0(\xi); \qquad G(\xi) = 1 - \xi K_1(\xi), \tag{4}$$

where K_0 and K_1 are the modified Bessel functions of the second kind and of order zero and one."

Introducing instead of A the parameter u defined by $F(u) = 1 - 1/n$ we have, from (1), approximately for large n

$$(5) \qquad F(x_1) = 1/n\left(\frac{u}{-x_1}\right)^k, \quad 1 - F(x_n) = 1/n\left(\frac{u}{x_n}\right)^k, \quad x_1 \leq 0, x_n \geq 0, k > 0.$$

For the variable ξ in Elfving's theorem, we obtain asymptotically

$$(6) \qquad \xi_k/2 = u^k \rho^{-k}.$$

We attach a subscript k to ξ to show its dependence on k. The moments of ξ_k are obtained from a formula given by Watson ([3], p. 388) as

$$(7) \qquad \overline{\xi_k^l} = 2^l \Gamma^2(1 + l/2)$$

and all moments of this variate exist.

3. Estimate of parameters. From N sets, each of n observations, we pick out the largest and the smallest, $X_{n,\nu}$ and $X_{1,\nu}$. We subtract from each observed extreme the central value, m, of the $N \cdot n$ observations. If each $x_{n,\nu} = X_{n,\nu} - \mathrm{m} \geq 0$ and $x_{1,\nu} = X_{1,\nu} - \mathrm{m} \leq 0$ the sample size is large enough.

Define $\eta = 1/\rho$. The first two moments of η are, from (7),

$$(8) \qquad \bar{\eta} = \frac{1}{u}\Gamma^2(1 + 1/2k), \qquad \overline{\eta^2} = \frac{1}{u^2}\Gamma^2(1 + 1/k).$$

Elimination of the parameter u from these two equations leads to

$$\frac{\overline{\eta^2}}{\bar{\eta}^2} = \frac{\Gamma^2(1 + 1/k)}{\Gamma^4(1 + 1/2k)}.$$

In terms of the coefficient of variation, V, this equation becomes

$$(9) \qquad \sqrt{1 + V^2} = \Gamma(1 + 1/k)/\Gamma^2(1 + 1/2k).$$

Substituting the value of V computed from the observations, we obtain an estimate of k, and hence can obtain an estimate of u from (8). This procedure is facilitated by Table 1.

4. The distribution of the geometric range. From a practical standpoint the geometric range itself is preferable to its reciprocal since it is easier to interpret and easier to calculate from the observed extremes. We want to establish its distribution $g_1(\rho)$. From the relation (6) of ρ to ξ_k and the knowledge of the distribution (4) of ξ_k we find

$$(10) \qquad G_1(\rho) = 1 - G(\xi_k) = 2u^k\rho^{-k}K_1(2u^k\rho^{-k})$$

and

$$(11) \qquad g_1(\rho) = \frac{2\xi_k\, k u^k}{\rho^{k+1}}K_0(\xi_k) = \frac{4k}{u}\left(\frac{u}{\rho}\right)^{2k+1} K_0\left(\frac{2u^k}{\rho^k}\right).$$

Since tables of these Bessel functions are available [2], the various probabilities and densities may be evaluated.

The simplest way to compare geometric ranges to the theory is the use of a probability paper (Figure 1). For its construction, consider the linear relation

$$\log \rho = \log u + (\log 2)/k - (\log \xi_k)/k \tag{12}$$

obtained from (6). Consequently we plot $-\log \xi_k$ on the abcissa and write the corresponding values $G_1(\rho)$, formula (10), on a horizontal axis. An upper parallel to the abscissa shows the return periods. The observed geometric ranges are plotted on the ordinate in a logarithmic scale. If the theory holds, the observed geometric ranges should be scattered about the straight line (12).

TABLE 1

The order k and the variation V of the reciprocal of the geometric range

Reciprocal Order $1/k$	Coefficient of variation V	Reciprocal Order $1/k$	Coefficient of variation V
.10	.088	.70	.556
.12	.104	.80	.632
.16	.138	.90	.709
.20	.171	.98	.772
.30	.251	1.00	.788
.40	.332	2.00	1.73
.50	.404	4.00	5.92
.60	.480	6.00	20.0

If less accurate estimates of u and k than those obtainable by the systematic methods (8) and (9), or the probability paper, will suffice, quick estimates can be obtained from the quantiles of the sample of geometric ranges. To the value $\rho = u$ corresponds, according to (6), $\xi_k = 2$ whence, from the tables [2], $G_1(u) = 2K_1(2) = .27973$. From N observed geometric ranges arranged in increasing magnitude we thus may pick out the mth, ρ_m, with the rank $m = .28\,N$ and use it as an estimate $u = \rho_m$. For the medians $\tilde{\xi}_k$ and $\tilde{\rho}$ we get $\tilde{\xi}_k = 1.257$ from the tables, and thus, by (6), $\tilde{\rho}^k = 1.591\,u^k$. This formula provides a quick estimate of k. We pick out the median $\tilde{\rho}$ of the N observed geometric ranges. Since we have an estimate of u, we obtain an estimate of k from

$$\frac{1}{k} = \frac{\log \tilde{\rho} - \log u}{\log 1.591} = 4.960 \log [\tilde{\rho}/\rho_m]. \tag{13}$$

5. Analogy between the geometric range and the range. A study of the various characteristics of the geometric range for distributions of Cauchy's type reveals structural similarities to the range for distributions of the exponential type.

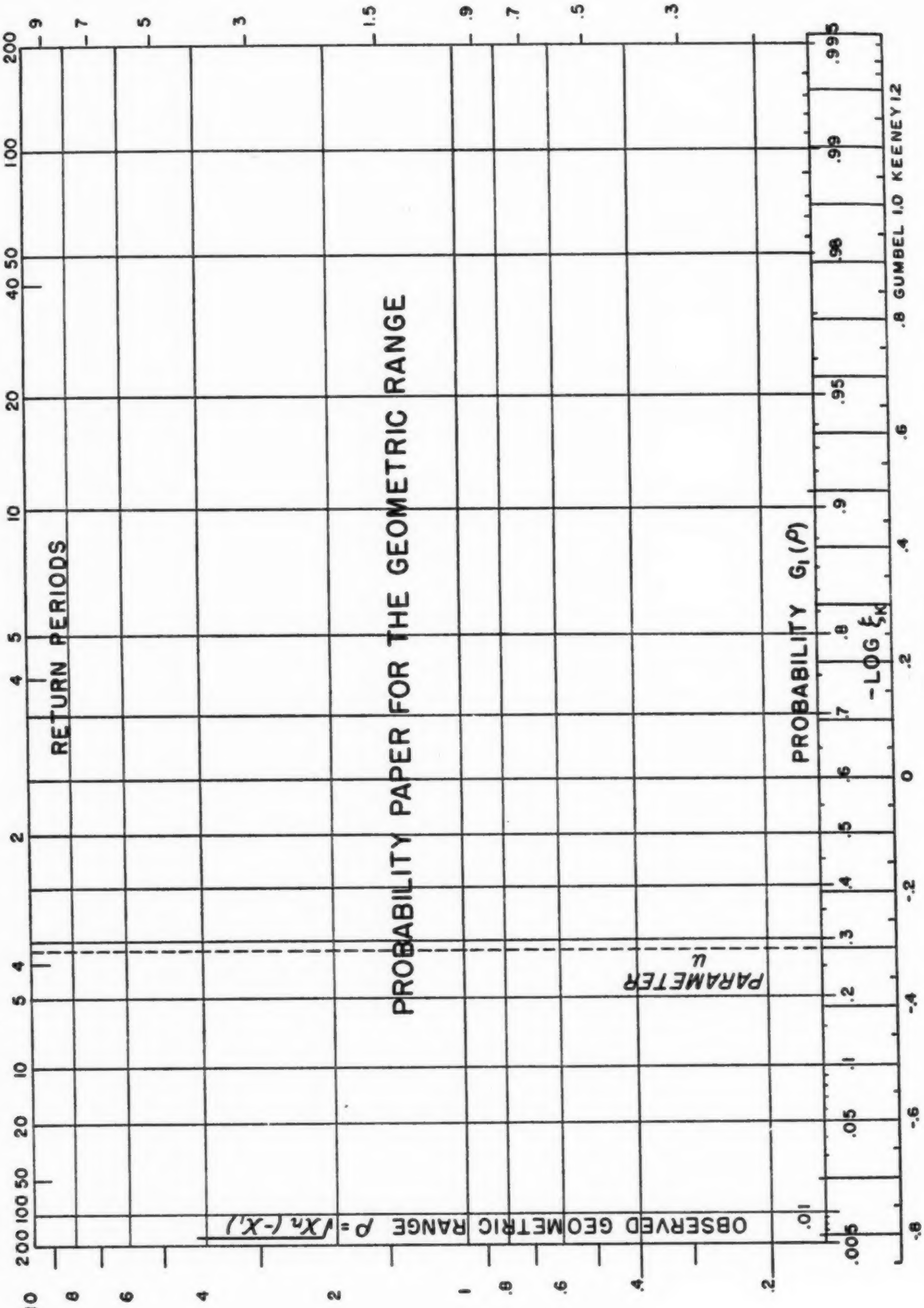
PROBABILITY PAPER FOR THE GEOMETRIC RANGE
RETURN PERIODS
PROBABILITY $G_1(\rho)$
$-\log \xi_K$
PARAMETER u
OBSERVED GEOMETRIC RANGE $\rho = \sqrt{X_n(-X_1)}$
GUMBEL 1.0 KEENEY 1.2

FIGURE 1

This is not altogether surprising, since (as shown in Table 2) after the appropriate transformations the probabilities of both are identical functions of the respective transformed variates.

Of course the two systems are mutually exclusive: if the observed ranges can be reproduced by the first system we conclude that all moments in the initial distribution exist. If on the other hand, the observed geometric ranges can be represented by the second sytem we conclude that no moments of an order greater than k exist.

TABLE 2

RANGES AND GEOMETRIC RANGES

Type of Initial Distribution	Exponential	Gauchy
Variate	Range	Geometric Range
Definition	$w = x_n + (-x_1)$	$\rho = \sqrt{x_n(-x_1)}$
Transformation	$z = 2\exp\left[-\frac{\alpha}{2}(x_n - x_1 - 2u)\right]$	$\xi_k = 2u^k\rho^{-k}$
Logarithm	$\lg z = \lg 2 - \frac{\alpha}{2}(x_n - x_1 - 2u)$	$\lg \xi_k = \lg 2 - \frac{k}{2}(\lg x_n + \lg(-x_1) - 2\lg u)$
Probability	$G(w) = z\,K_1(z)$	$G_1(\rho) = \xi_k\,K_1(\xi_k)$
Distribution	$g(w) = \frac{\alpha z^2}{2}K_0(z)$	$g_1(\rho) = \frac{4k}{u}\left(\frac{\xi_k}{2}\right)^{2k+1}K_0(\xi_k)$
Median	$\tilde{w} = 2u + .9286/\alpha$	$2\lg\tilde{\rho} = 2\lg u + .9286/k$
Mean	$\bar{w} = 2u + 2\gamma/\alpha$	$\lg\overline{\rho^{-1}} = -\lg u + 2\lg\Gamma(1 + \frac{1}{2}k)$

REFERENCES

[1] G. Elfving, "The asymptotical distribution of range in samples from a normal population," *Biometrika*, Vol. 35 (1947).

[2] *Tables of the Bessel-functions*, Vol. 6, British Association for the Advancement of Science, Cambridge, 1937.

[3] G. N. Watson, *Theory of Bessel-functions*, Cambridge University Press, 1944.

FIGURE 1

REMARK ON W. M. KINCAID'S "NOTE ON THE ERROR IN INTERPOLATION OF A FUNCTION OF TWO INDEPENDENT VARIABLES"

By T. N. E. Greville

Federal Security Agency

In a review of Dr. W. M. Kincaid's "Note on the Error in Interpolation of a Function of Two Independent Variables," (*Annals of Math. Stat.*, Vol. 19 (1948),

pp. 85–88) which appeared in *Mathematical Reviews*, Vol. 9 (1948), p. 470, I stated that "a more simple and elegant, and equally general, expression is obtainable by a simple adaptation of formula (41), p. 215, of J. F. Steffensen's book, *Interpolation*."

This statement is not entirely correct and is also misleading in its implications since Dr. Kincaid's expressions are actually more general in certain respects, and simplicity and generality are not the only considerations nor, in this case, the most important ones. In setting up an expression for the remainder in an interpolation formula, the primary objective is to secure an efficient appraisal of the remainder. In this respect, Dr. Kincaid's expressions are superior as they involve only the higher derivatives of the function it is desired to represent, whereas Steffensen's method would always involve a first derivative term in such a way as to prevent any refinement of estimates of the error by introducing additional given values.

REMARK ON MY PAPER "ON A THEOREM OF HSU AND ROBBINS"

By P. Erdös

Syracuse University

Professor Robbins kindly pointed out that in my paper mentioned in the title (*Annals of Math. Stat.*, Vol. 20 (1949), p. 286–291) I have misquoted a statement in the paper of Hsu and Robbins ("Complete Convergence and the Law of Large Numbers" *Proc. Nat. Acad. of Sci.*, Vol. 33 (1947), p. 25–31). I attribute to Hsu and Robbins the conjecture (notations of my paper) that if $\sum_{n=1}^{\infty} M_n < \infty$ then (1) and (2) hold, and proceed to give a counter example. However, the conjecture of Hsu and Robbins is not the above false one but the following: If $\sum_{n=1}^{\infty} M_n < \infty$ and (1) holds then (2) also holds. This conjecture is true and is in fact proved in my paper.

Professor Robbins also points out that a slight modification of my theorem can be stated in a more concise form as follows: Let $X_1, X_2, \cdots$ be a sequence of independent random variables having the same distribution function $F(x)$, and let

$$Y_n = (1/n)\ (X_1 + \cdots + X_n)$$

Then the necessary and sufficient condition that

$$\sum_{n=1}^{\infty} P_r\{|Y_n| > \epsilon\} < \infty, \qquad \text{for every } \epsilon > 0,$$

is that

$$\int_{-\infty}^{\infty} x\, dF(x) = 0, \qquad \int_{-\infty}^{\infty} x^2\, dF(x) < \infty.$$

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the New York meeting of the Institute, December 27–30, 1949)

1. **The Asymptotic Distribution of the Extremal Quotient.** E. J. Gumbel, New York, and R. D. Keeney, Metropolitan Life Insurance Company, New York.

The extremal quotient is the ratio of the largest to the absolute value of the smallest observation. Its analytical properties for symmetrical, continuous and unlimited distributions are obtained from a study of the auto-quotient defined as the ratio of two non-negative variates with identical distributions. The relation of the two statistics is established by proving that, for sufficiently large samples from an initial distribution with median zero, the largest (or smallest) value may be assumed to be positive (or negative) and that the extremes are independent. The logarithm of the extremal quotient has asymptotically a symmetrical distribution. Its median is unity. As many moments exist for the extremal quotient as moments and reciprocal moments exist simultaneously for the initial variate. For the exponential type of initial distributions, the asymptotic distribution of the extremal quotient can only be expressed by a complicated integral which may be approximated in the interval $\frac{1}{2} < q < 2$ by the logarithmically transformed normal probability function. In this case, no moments exist. For the Cauchy type, the asymptotic distribution of the extremal quotient is very simple. The logarithm of the extremal quotient has the same (logistic) distribution as the midrange for initial distributions of exponential type. For both initial types, the asymptotic distributions of the extremal quotients possess one parameter which may be estimated from the observations.

2. **A Second Formula for Partial Sums of Hypergeometric Series having the Unit as Fourth Argument.** Hermann von Schelling, Naval Medical Research Laboratory, U. S. Submarine Base, New London, Conn.

If the arguments α and β are changed after the summation, published *Ann. Math. Stat.* Vol. *20,* (1949) p. 120, and this method is applied a second time, a new formula results for partial sums of $F(\alpha,\beta,\gamma;1)$. A simple recurrence formula is developed for these partial sums. The new equation is a numerical short cut as it is demonstrated with an example.

3. **A Coverage Distribution.** Herbert Solomon, Office of Naval Research, Washington, D. C.

Consider a fixed target circle of radius T_R and center at a distance R from an aiming point. Let N circles each of radius W_R be dropped at the aiming point with their centers subject to a bivariate normal distribution with circular symmetry, the common standard deviation denoted by σ. Define γ as the set theoretical sum of the N random circles with the fixed circle and let c be the ratio of γ to the total area of the fixed circle. Then it is desired to find P_{c_o} where

$$P_{c_o} = P\{c \geq c_o \mid T_R, W_R, R, N\}$$

where T_R, W_R, and R are in σ units. Define $R^* = W_R + aT_R$ where $a = a(c, W_R, T_R)$; $|a| \leq 1$. It is shown that for $N = 1$, the family of curves in the RR^* plane defined by P_{c_o} = constant have a slope, m, given by

$$m = \frac{I_1(RR^*)}{I_0(RR^*)}$$

where I_k is the modified Bessel Function of k^{th} order. In fact as the product

RR^* approaches infinity, m approaches unity. From these results, the contours of equal probability are easily determined. When $N > 1$, overlap considerations make the computation of explicit values for P_{c_0} intractable. However, in this case, upper and lower bounds for P_{c_0} can be obtained.

4. The Problem of the Greater Mean. R. R. BAHADUR AND HERBERT ROBBINS, University of North Carolina, Chapel Hill.

"Optimum" solutions (in the sense of Wald's theory of statistical decision functions) are obtained for the "problem of the greater mean". Let π_i $(i = 1,2)$ be normal populations with means m_i and common variance σ^2, all unknown, and denote the arbitrary but given set of possible parameter points $\omega = (m_1, m_2 : \sigma)$ by Ω. Suppose that a set of $n_1 + n_2$ independent observations is drawn, n_i from π_i, and let $v = (x_{11}, \cdots, x_{1n_1}; x_{21}, \cdots, x_{2n_2})$ denote the sample point. Any measurable function $f(v)$ such that $0 \leq f(v) \leq 1$ is called a decision function. Given a "risk function" $r(f \mid \omega)$ defined for all f and all $\omega \epsilon \Omega$, a decision function $f^*(v)$ is "optimal" if (i) $\sup[r(f^* \mid \omega)] = \inf \sup [r(f \mid \omega)]$, and (ii) no decision function is "uniformly better" than $f^*(v)$. If $f^*(v)$ is the unique (up to sets of measure 0) decision function with property (i), it is "optimum". *Case* 1. Given any decision function $f(v)$ and any $\omega \epsilon \Omega$, let

$$r(f \mid \omega) = \max [m_1, m_2] - m_1 E [f \mid \omega] - m_2 E[1 - f \mid \omega].$$

Let

$$f^\circ(v) = \begin{matrix} 1 \text{ if } \bar{x}_1 > \bar{x}_2 \\ 0 \text{ otherwise} \end{matrix} \qquad \left(\bar{x}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}\right)$$

It is shown that under certain conditions on Ω, $f^\circ(v)$ is optimum. *Case* 2. Given any decision function which takes on only the values 0 and 1, corresponding to the two decisions "$m_1 \leq m_2$" and "$m_2 \leq m_1$" respectively, and any $\omega \epsilon \Omega$, let

$$r(f \mid \omega) = P(\text{incorrect decision} \mid \omega, f).$$

It is shown that under certain conditions on Ω, $f^\circ(v)$ is optimal. The conditions on Ω are very similar in the two cases, and are likely to be satisfied in most applications. However, it is shown by examples that there exist non-degenerate types of Ω with respect to which decision functions other than $f^\circ(v)$ are *uniformly better* than $f^\circ(v)$. The methods of the paper can be applied to a number of similar problems.

5. Some Extensions of Bayes' Theorem. F. C. Leone, Case Institute of Technology, Cleveland 6, Ohio.

There is some past or *a priori* knowledge about the quality of a population of lots and a sample is taken from a random lot. What can be said about the lot from which this sample is taken? We are incorporating the results of our experiment or sample with the previous knowledge to form a judgment. From the *a priori* distribution and a sample of n with c defectives, say two in twenty-five, we form an *a posteriori* distribution of all two in twenty-five cases. From this distribution we can answer questions such as: "What is the *a posteriori* probability that a lot producing a two in twenty-five result should have a proportion of defectives ten per cent or below?" We consider as our *a priori* situation such distributions as the rectangular, triangular, normal, Pearson's Type III and Type I. These extensions are applied to some industrial data. In considering lot quality on one hundred per cent inspection, the *a priori* distributions of these data are mostly U-shaped with some bell-shaped and J-shaped. In some cases a Pearson Type I proves to be a good fit for the *a priori* distribution.

6. **On Optimum Selections from Multinormal Populations.** Z. W. BIRNBAUM AND D. G. CHAPMAN, University of Washington, Seattle.

Let $(X, Y_1, \cdots, Y_n)$ have an $(n + 1)$-dimensional non-singular normal probability density $f(X, Y_1, \cdots, Y_n)$. By "selection" in $(Y_1, \cdots, Y_n)$ we shall understand a measurable function $\varphi(Y_1, \cdots, Y_n)$ such that $0 \leq \varphi \leq 1$ for all $Y_1, \cdots, Y_n$. By a "truncation in $(Y_1, \cdots, Y_n)$ to the set Ω" we understand a selection $\varphi(Y_1, \cdots, Y_n)$ such that $\varphi = 1$ for $(Y_1, \cdots, Y_n)$ in Ω, and $\varphi = 0$ in Ω. A "linear truncation" will be a truncation to a set defined by a condition of the form $\sum_{c=1}^{n} c_i Y_i \geq k$. Using a slight generalization of Neyman-Pearson's fundamental lemma, the following theorems are proven: among selections for which the expectation of X, after selection, assumes a fixed value, the one which maximizes the "retained" portion of the universe $\int \cdots \int \varphi(Y_1, \cdots, Y_n) f(X, Y_1, \cdots Y_n) dX dY_1 \cdots dY_n$ is a linear truncation. Among all the selections for which a given quantile of X, after selection, assumes a fixed value, the one which maximizes the retained portion of the universe is a linear truncation. (Research under the sponsorship of the Office of Naval Research).

7. **Simple Regression Analysis with Autocorrelated Disturbances.** HOWARD L. JONES, Illinois Bell Telephone Company, Chicago.

When the disturbances in a regression equation are connected by a linear difference equation, the parameters of both equations can be estimated simultaneously by maximizing a function that describes the joint probability of the disturbances or a linear function thereof. This note discusses a simple example.

8. **A Test of Klein's Model III for Changes of Structure.** A. W. MARSHALL, The Rand Corporation, Santa Monica, Calif.

This paper suggests a test of equations from linear stochastic equation systems on the basis of observations not included in the original computation period. Rejection regions of approximately the right size (asymptotically correct) are constructed and the use of naive economic models as an auxillary test are suggested. The procedure is applied to Klein's Model III, the results are tabulated and discussed.

9. **An Application of the Theory of Extreme Values to Economic Problems.** S. B. LITTAUER, Columbia University, AND E. J. GUMBEL, New York.

Most studies of economic time series have been concerned with establishing regularities of behavior, often by analogy with mechanical systems. Much as regularity in economic phenomena is desirable, such evidence as has been available leaves the reality of this sought for regularity considerably in doubt. It seems more fruitful rather to ask the question, "What is the pattern of the non-regularity" and if reasonably answered, to offer some verifiable form of explanation therefor. It seems further desirable that any attempt at "scientific" explanation of economic phenomena be fortified by evidence of statistical stability supported by criteria such as were established by Shewhart for the control of quality of manufactured product. In the present instance certain concepts of experimental inference, which seem natural therefor, are employed in order to give some general and plausible unity to the behavior of economic time series.

Following upon the postulates of the theory presented here, the appropriate formal development employs concepts of statistical quality control and of the statistical theory of extreme values. Within this theory the importance of the absence of statistical stability

is emphasized, and the relevance of the use of concepts in extreme values is made evident. By introducing a superuniverse, peaks and troughs are random expressions of a super chance-"cause" system. The use of these statistical concepts is not motivated by mere analogy but rather as the natural means for explanation of the phenomena studied.

A number of examples of the application of these statistical methods to selected series are offered as evidence of the workability of the theory here presented. The extremes of the Dow-Jones index of selected industrials show that the 1928 value was completely outside the previous levels and should not have been considered as a "stable high plateau basic for perpetual prosperity". Instead this should have suggested the imminent breakdown. The validity of the application of the theory of extreme values to these phenomena is not so strongly substantiated as are the many applications that have been made of them to flood frequencies, wind velocities, extreme temperatures, breaking strengths and other natural phenomena. Nevertheless the results here obtained are highly suggestive of a tenable economic hypothesis.

10. Bias Due to the Omission of Independent Variables in Ordinary Multiple Regression Analysis. (Preliminary Report). T. A. Bancroft, Iowa State College, Ames.

Given n observations of the dependent variate y and the independent variates x_1, $x_2, \cdots, x_k, \cdots, x_r$, $k < r$, all variates measured from their respective sample means, and we have calculated the ordinary regression of y on the first k variates and y on all r variates. We define ordinary multiple regression as the single-equation approach, error only in y which is assumed normally and independently distributed with zero mean and variance σ^2, the x_i being fixed from sample to sample.

In order to determine whether to omit or retain the last $(r - k)$ independent variates we formulate a rule of procedure: calculate Snedecor's $F =$

$$\frac{\text{Reduction in } Sy^2 \text{ due to } (r-k) \text{ variates}/(r-k)}{\text{Error mean square after fitting all } r \text{ variates}}$$

If F is non-significant at some assigned significance level α, we pool the sums of squares and degrees of freedom, involved in the numerator and denominator of F, to obtain an estimate of the error σ^2, and fit y on the first k variates only. If F is significant at the assigned significance level we use the denominator only in F for our estimate of σ^2 and hence fit y on all r variates.

The object of this investigation is to determine the bias in our estimate e^* of σ^2, if we follow such a rule of procedure. The bias turns out to be

$$\frac{2\sigma^2\lambda}{n_1+n_2} + \sigma^2 e^{-\lambda} \sum_{i=0}^{\infty} \left[I_{x_0}\left(\frac{n_2}{2}+1, \frac{n_1}{2}+i\right) - I_{x_0}\left(\frac{n_2}{2}, \frac{n_1}{2}+i\right) \frac{-2i}{n_1+n_2} I_{x_0}\left(\frac{n_2}{2}, \frac{n_1}{2}+i\right)\right] \frac{\lambda^i}{i!}$$

where

$$x_0 = \frac{n_2}{n_2 + n_1\alpha}, \qquad \lambda = \frac{\sum_{i=k+1}^{r} (\beta_i')^2}{2\sigma^2},$$

n_1 and n_2 are the respective degrees of freedom for the numerator and denominator of F, and $\sum_{i=k+1}^{r} (\beta_i')^2$ is a function of the population regression coefficients $\beta_{k+1}, \cdots, \beta_r$. The bias is discussed for selected values of the parameters involved.

11. **Estimating Parameters of Pearson Type III Populations From Truncated Samples.** A. C. COHEN, JR., The University of Georgia, Athens.

The method of moments is employed with 'single' truncated random samples (1) to estimate the mean, μ, and the standard deviation, σ, of a Pearson Type III population when α_3 is known and (2) to estimate μ, σ, and α_3 when only the form of the distribution is known in advance. No information is assumed to be available about the number of variates in the omitted portion of the sample. The results obtained can be readily applied to practical problems with the aid of "Salvosa's Tables of Pearson's Type III Function." An illustrative example is included in the paper.

12. **The Cyclical Normal Distribution.** E. J. GUMBEL, New York.

The usual normal distribution becomes invalid for variates, like an angle, lying on the circumference of a circle. The distribution of such variates was established by R. von Mises by the same methods as used for the classical derivation. The cyclical normal distribution is symmetrical about a mode and antimode. The probability function is proportional to an incomplete Bessel function of the first kind and of order zero for an imaginary argument, and contains two parameters, the direction of the resultant vector and a parameter k linked to the absolute amount of the vector. The parameters may be estimated by the method of maximum likelihood. For $k = 0$, the distribution degenerates into a uniform cyclical distribution. If k is of the order 3, the distribution approaches the linear normal one, k being the reciprocal of the variance. With increasing values of k, the distribution looses its cyclical character and becomes concentrated in a narrow strip. This distribution holds for symmetrical unimodal values varying according to pure chance about a unique mode in a closed space (as the angles of the wind directions) or a closed time, and gives a theoretical model for the variations of temperatures, pressures, rainfalls, storms, discharges, floods, death- and birth rates over the year, and earth quakes over the day. The comparison between theory and observations in plotting the square roots of the frequency on polar coordinate paper provides a statistical criterion for the regularity of cyclical phenomena. (Work done in part under contract *W* 44/109/*QM*/2202 with the Research and Development Branch, Office of the Quartermaster General).

13. **Treatment of Attenuation Problems by Random Sampling.** H. KAHN AND T. HARRIS, The Rand Corporation, Santa Monica, Calif.

Exact analytical calculations of the transmission of energy by particles through shields are difficult; to avoid them random sampling methods may be resorted to. The straightforward procedure of simulating life histories of particles, using random number tables, may be used for thin shields, but in the case of thick shields with tremendous attenuations, tremendous numbers of particles would be required. In order to obtain reasonably small standard errors, using reasonable numbers of simulated life histories, it is necessary to modify the original problem to one having a lower attenuation factor, the solution bearing a known relation to the solution of the original problem. Alternatively, this may often be regarded as an application of well known statistical sampling procedures, such as representative sampling or importance sampling. Various special procedures can be devised. One of the first was the splitting technique due to J. v. Neumann. Among others may be mentioned the exponential transformation, a simple analytic transformation of the original problem into one having a much lower attenuation factor.

14. **On the Existence of Nearly Locally Best Unbiased Estimates.** HERMAN RUBIN, Stanford University, Stanford, Calif.

For any family $\mathcal{F}$ of distributions, and any distribution F_0 of $\mathcal{F}$, there exists a bilinear function K whose arguments are all parameters defined for all distributions of $\mathcal{F}$ and for

which there exist unbiased estimates which have finite variance if F_0 is the true distribution, and which has the following properties: (1) If θ is any parameter in the domain of K, and t is any unbiased estimate of θ, then $var(t \mid F_0) \geq K(\theta, \theta)$. (2) This result is best possible, i. e., for any θ there is an unbiased estimate t of θ whose variance differs from $K(\theta, \theta)$ by less than any preassigned amount.

15. The Experimental Evaluation of Multiple Definite Integrals. George W. Taylor, U. S. Army Electronics Laboratory, San Diego, Calif.

When one is forming an estimate of the total, or mean value, of some quantity, sampling at carefully selected points will frequently be preferable to employing a method which involves randomization. The estimation of the total volume of water in a given lake or the amount of energy being released in a given time and space, are examples of problems where specified points for sampling should result in a reduction in the error of estimate. These and similar problems lead naturally to numerical integration methods. In the case of single integrals, Gauss' and Tchebychef's formulae yield maximum efficiency with respect to controlling the polynomial error and statistical error respectively, but often the Newton-Cotes formulae can be applied more conveniently.

For the evaluation of double integrals, an eight point and a thirteen point formula for fifth degree accuracy and a twelve point and a twenty-one point formula for seventh degree accuracy have been developed for integrating over a rectangle and similar formulae have been developed for integrating over areas bounded by a parabola and a straight line or by two parabolas. The following system of equations is employed in developing these formulae:

$$\sum_{\alpha=1}^{m} R_\alpha x_\alpha^i y_\alpha^j = C_{ij}, \quad \text{for all } i, j \text{ for which } i + j \leq 2n,$$

$$\text{and where } C_{ij} = \frac{a^{i'} b^j}{(i+1)(j+1)} \text{ for both } i \text{ and } j \text{ even,}$$

$$= 0 \text{ otherwise.}$$

Formulae for the numerical evaluation of triple integrals taken over a rectangular parallelopiped are developed, including a twenty-one point formula with fifth degree accuracy. It is shown that comparable formulae can be developed for integrating functions of more than three variables and a $2n + 1$ point formula with third degree accuracy for integrating a function of n variables over a rectangular n-space is obtained.

16. Tests of Fit of a Cumulative Distribution Function over Partial Range of Sample Data. Bradford F. Kimball, New York State Dept. of Public Service, New York.

Case 1. Sample data are completely ordered over range tested.

Let the $n + 1$ true frequency differences associated with an ordered random sample of n values of x be denoted by u_i. The *cdf* of a theoretical test function based on m of the above frequency differences is identified and methods of approximating it are discussed.

Case 2. Sample data in k ordered groups over range tested.

Let $\Delta_i F$ denote the true frequency differences over the k sample intervals to be covered by the test. Let m_i denote the number of unit frequency differences u_i covered by the *ith* interval. Define M and W by

$$M + 1 = \sum_k m_i, \qquad M \leqq n;$$

$$W = \sum_k \Delta_i F, \qquad W \leqq 1.$$

A theoretical function Z is defined by

$$Z = \frac{(M+1)(M+2)}{k-1} \sum_{k} \frac{[\Delta_i F - m_i W/(M+1)]^2}{m_i}.$$

Set

$$Y = Z/W^2.$$

The *cdf* of Y is identified and methods of approximation to it are discussed.

Applications to testing agreement of sample with hypothetical *cdf* of universe are considered for both cases in some detail.

17. Large Sample Tests for Comparing Percentage Points of Two Arbitrary Continuous Populations. A. W. Marshall and J. E. Walsh, The Rand Corporation Santa Monica, Calif.

Let us consider two continuous populations, the first with density function $f(x)$ and $100\alpha\%$ point θ_α, the second with density function $g(x)$ and $100\beta\%$ point ϕ_β. These two populations are arbitrary except that $f(\theta_\alpha) \neq 0$, $g(\phi_\beta) \neq 0$ and both $f'(\theta_\alpha)$, $g'(\phi_\beta)$ exist and are continuous in the vicinity of the specified points. This paper presents significance tests for $\theta_\alpha - \phi_\beta$ which are based on large samples from these populations. The exact significance level of a test is not known but its value is bounded within reasonably close limits (asymptotically). Efficiency properties of these tests (compared to the corresponding noncentral t-tests) are investigated for the case in which both populations are normal and the ratio of variances is known. Results are also derived for simultaneously testing $\theta_\alpha - \phi_\beta$ and $f(\theta_\alpha)/g(\phi_\beta)$. These tests have known significance levels (asymptotically). A particular application of tests of this type occurs when it is desired to test whether two samples came from the same population and agreement of the two populations in a specified region is to be emphasized. For this special case, the significance levels of the resulting tests are reasonably accurate for moderate as well as large sized samples.

18. On the Distribution of Wald's Classification Statistic. H. L. Harter, Michigan State College, East Lansing.

A study is made of the distribution of the classification statistic introduced by Wald. The exact distribution of V in the univariate case, as obtained by the use of characteristic functions and contour integration, is given for both degenerate and non-degenerate cases. The problem of classifying an individual into one or the other of two populations, using the statistic V, is discussed. In the multivariate case, examples are given of the distribution of an approximation to V suggested by Wald. The procedure here consists integrating out two variables from the joint distribution of three variables to find the distribution of the third. Four cases arise, depending upon whether the sample size and the number of variates are even or odd. Since this approximation is valid only for large samples, an attempt is made to find an approximation which is asymptotically equivalent to it as the sample size increases, but which is valid also for small samples. Results are given for a sampling experiment performed to determine an empirical distribution of V for a specific small sampling case, using a population of 10,000 pieces modeled after Shewhart's normal bowl. Obstacles in the path of practical applications are discussed.

19. Analysis of Extreme Values. W. J. Dixon, University of Oregon, Eugene.

Consider a population $N(\mu, \sigma^2)$ contaminated by introducing a certain proportion of values from a population $N(\mu + \lambda\sigma, \sigma^2)$ or $N(\mu, \lambda^2\sigma^2)$. The performance of various statistics for discovering these contaminators is assessed by sampling methods for samples of size 5 and 15. (This research was sponsored by the Office of Naval Research).

20. A Note On The Variance Of Truncated Normal Distributions. A. C. Cohen, Jr., The University of Georgia, Athens.

Formulas are derived whereby the variance of truncated normal distributions can readily be computed with the aid of an ordinary table of areas and ordinates of the normal frequency function. These results are applicable to certain tolerance problems involved in Statistical Quality Control. Their use will enable one to make computations required in solving such problems without resorting to Karl Pearson's relatively inaccessible tables of "Values of the Incomplete Normal Moment Functions".

21. Some Estimates and Tests Based on the r Smallest Values in a Sample (By Title). J. E. Walsh, The Rand Corporation, Santa Monica, Calif.

Let us consider a situation where only the r smallest values of sample of size n are available. This paper investigates the case where n is large and r is of the form $pn + 0(\sqrt{n})$. Properties of some well known estimates and tests of the $100p\%$ population point (based on statistics of the type used for the sign test) are investigated. If the sample is from a normal population, these nonparametric results have high efficiencies for small values of p (at least 95% if $p \leq 1/10$). The other investigations are restricted to the case of a normal population. Asymptotically "best" estimates and tests of the population percentage points are derived for the case where the population variance is known. If the population variance is unknown, asymptotically most efficient estimates and tests can be obtained for the smaller population percentage points by suitable choices of p and $0(\sqrt{n})$. The results of the paper have application in the field of life testing. There the r smallest sample values can be obtained without the necessity of obtaining the remaining sample values. By starting with a larger number of units but stopping the experiment when only a small percentage have "died", it is often possible to obtain the same amount of "information" with a substantial saving in cost and time over that required by starting with a smaller number of units but continuing until all have "died".

22. Some Comments on the Efficiency of Significance Tests (By Title) J. E. Walsh, The Rand Corporation, Santa Monica, Calif.

A method sometimes used to measure the efficiency of a significance test consists in associating a statistic with the test and defining the efficiency of the test to be the efficiency of this statistic considered as an estimate. This paper investigates the power function implications of this method of defining the efficiency of a test. Examples are presented which show that an estimate efficiency of $100E\%$ does not necessarily imply that the corresponding most powerful test based on $100E\%$ as many sample values has approximately the same power function as the given test (for the admissible set of alternative hypotheses). In several of the examples it was found that estimate efficiency makes no allowance for the effect of significance level while the relationship between the power functions of the given test and the corresponding most powerful test changes noticeably with respect to significance level. Some of these examples are non-asymptotic while others are asymptotic. However, results are obtained for the asymptotic case which indicate that this equality of power functions does hold for a rather broad class of significance tests if the pertinent statistics have distributions which are asymptotically normal.

23. Application of Sequential Sampling Method to Check the Accuracy of a Perpetual Inventory Record. Joseph B. Jeming, New York.

The problem of checking the continuing property records of a large utility company is handled by an application of the sequential sampling method as developed by the Statistical Research Group,

Columbia University. Without the application of a sampling procedure the problem can only be solved either by a complete physical inventory which is very costly, or by a cycle check which takes many years to complete. By use of the sequential sampling method, results of desired accuracy are obtained quickly and at very low cost since an extremely small percentage of field inspection for the mass property accounts of any large utility produces satisfactory conclusions.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest.

Personal Items

Dr. Ralph A. Bradley accepted an appointment as Assistant Professor in the Mathematics Department of McGill University, Montreal, Canada after receiving his Ph.D. in mathematical statistics at the University of North Carolina in June, 1949.

Mr. Fred J. Clark, Jr. received his master of science degree in mathematics from the University of Illinois in August, 1949 and is now employed by the University of California at the Sandia Laboratory in Albuquerque, New Mexico.

Professor J. L. Doob is on leave from the University of Illinois to teach at Cornell University for the academic year 1949–1950.

Mark W. Eudey obtained his Ph.D. degree in statistics at the University of California, Berkeley, and is now Vice President of California Municipal Statistics, Inc.

Dr. Joseph L. Hodges, Jr. has been promoted to Assistant Professor and Research Associate at the Statistical Laboratory, University of California, Berkeley.

Professor Paul Horst, formerly of the Department of Psychology, University of Washington, is now Director of Research at the Educational Testing Service, Princeton, New Jersey.

Dr. Fred C. Leone, formerly an Instructor and a Research Fellow at Purdue University, has been appointed Instructor in the Mathematics Department and Director of the Statistical Laboratory at the Case Institute of Technology.

Mr. Fred W. Lott, who has been studying at the University of Michigan for his Ph.D., has accepted an assistant professorship at Iowa State Teachers College, Cedar Falls, Iowa.

Dr. Francis McIntyre has resigned as Director of Export Control, Office of International Trade, U. S. Department of Commerce, Washington, D. C. to accept a post as Director of Economic Research, California Texas Oil Co., 551 Fifth Avenue, New York, New York.

Mr. R. B. Murphy, who has been a graduate student at Princeton University has accepted an instructorship in the Mathematics Department of Carnegie Institute of Technology.

Professor Jerzy Neyman, Director of the Statistical Laboratory, University of California at Berkeley, will be on sabbatical leave for the Spring Semester, 1950.

Mr. Monroe L. Norden, formerly of the Glenn L. Martin Co., is now a Mathematical Statistician with the Operations Research Office, Johns Hopkins University, Ft. Lesley, J. McNair, Washington 25, D. C.

Mr. D. Martin Sandelius, formerly a Research Assistant in the Institute of Statistics, Uppsala, Sweden, has been appointed Lecturer in the Mathematics Department, University of Washington, Seattle, for the academic year 1949–1950.

After completing his graduate work at Ohio State University, Dr. William J. Schull accepted a position with the Atomic Bomb Casualty Commission. He is now in Japan as a geneticist working on follow-up studies at Hiroshima.

Miss Elizabeth L. Scott obtained her Ph.D. degree in statistics at the University of California, Berkeley and was promoted to Lecturer and Research Associate at the Statistical Laboratory.

Miss Ester Seiden obtained her Ph.D. degree at the University of California, Berkeley and was promoted to Lecturer and Research Associate at the Statistical Laboratory.

Mr. Iriving H. Siegel is on leave from his position as Chief Economist at the Veterans Administration until June 30, 1950, to serve as Lecturer in Political Economy at the Johns Hopkins University and as a member of the Johns Hopkins University Operations Research Office staff.

Dr. Charles M. Stein, Assistant Professor and Research Associate at the Statistical Laboratory, University of California, Berkeley, will be on leave for the academic year 1949–1950 and will be working in Paris as a National Research Fellow.

Alfred James Lotka

Alfred James Lotka, a Fellow of the Institute, died in Red Bank, New Jersey, on December 5, 1949. He was born of American parents in Poland, March 2, 1880, and had his early schooling in France. His academic training was received at Birmingham, England (B.Sc., 1901, and D.Sc., 1912), Cornell (M.A., 1909), and Johns Hopkins (1922–1924). Dr. Lotka came to the Statistical Bureau of the Metropolitan Life Insurance Company in 1924 and retired as Assistant Statistician in 1947. His major contributions were his highly original work on the mathematical theory of evolution, on the mathematical analysis of population, and on the theory of self-renewing aggregates. Altogether, Dr. Lotka had almost 100 papers in these fields in technical and scientific journals, both here and abroad. The essentials of his work are summarized in his books, "The Elements of Human Biology" and "Theorie analytique des associations biologiques." He was, in addition, a joint author on several books in the field of public health.

Dr. Lotka was a past president of the American Statistical Association and of the Population Association of America. He had recently been active as American Vice-President of the International Union for the Study of Population.

Statistical Summer Session in Berkeley, Calif.

Following the established pattern, there will be held this year a Statistical Summer Session at the University of California, Berkeley. The faculty will include William G. Cochran of Johns Hopkins University, Benjamin Epstein of Wayne University, Erich L. Lehmann of the University of California, Paul Lévy

of the Ecole Polytechnique, Paris, France and Gottfried E. Noether of New York University.

Courses will be offered on both the graduate and the undergraduate levels. The graduate courses, all given during the First Summer Session, June 19 to July 29, are meant primarily for students who either have already obtained their Ph.D. degree or are working toward it. No specific prerequisites to graduate courses will be required. The graduate program includes (i) a course on design of experiments and a seminar on analysis of variance by W. G. Cochran, (ii) a course on theory of estimation by E. L. Lehmann, and (iii) a course and a seminar on random variables and random functions by Paul Lévy.

Inquiries should be addressed to the Office of the Summer Sessions, 1A Administration Building, University of California, Berkeley 4, California.

At a meeting of its Executive Council, AAPOR has laid plans for its 1950 meetings to be held jointly with the World Association for Public Opinion Research (WAPOR) at Lake Forest College, near Chicago, June 16 to 20.

The program which is now being planned will be designed to fit the needs of the Association's membership, which is composed of leaders in both the academic and commercial fields.

The Council of the Institute of Mathematical Statistics requested Professor Harold Hotelling to communicate to Professor S. S. Wilks its appreciation of his editorship of the *Annals* during the years 1938 to 1949. On the recommendation of the Council Professor Hotelling's letter is reproduced below.

January 6, 1950

Professor Samuel S. Wilks
Fine Hall
Princeton, New Jersey
Dear Professor Wilks:

In behalf of the Council of the Institute of Mathematical Statistics and by its direction, I write to express the appreciation we all feel for the splendid efforts which you have expended so freely upon the *Annals of Mathematical Statistics*, and which have been so conspicuously successful in establishing it as a sound and reputable journal. The years of your editorship are memorable ones for the history of statistics, and your contribution to making them so is of first importance.

Very sincerely,
Harold Hotelling

New Members

The following persons have been elected to membership in the Institute

(August 23, 1949 to November 30, 1949)

Anderson, Oskar, Ph.D. (Kiel) Professor, University of Munich, *Konigin-Strasse 69, Munich (Munchen), Germany*

Puente Arroyo, Felix Jorge, CPA, (Univ. Nal. Litoral) Professor titular Mathematics, *Italia 1550, Rosario, Republica Argentina.*

Arvanitis, Ernest A., A.B. (Boston Univ.) Student at Columbia University, *43-18 40th Street, Sunnyside, L. I., New York*

Bhatt, Narbheshanker M., Ph.D. (Edinburgh Univ.) Professor of Statistics, Commerce College, Behind Raopura Tower, Baroda, India

Bose, Raj Chandra, D. Litt (Calcutta Univ.) Professor of Mathematical Statistics, University of North Carolina, *110 Noble Street, Chapel Hill, North Carolina.*

Carreiro, Oscar Ediwaldo Porto, Civil Engineer (Univ. of Brazil) Professor da Faculadade de Ciencias Economicas, Avenida Sao Sebastiao 266, Sao Paulo, Brazil.

Crump, Phelps P., B.S. (Iowa State) Graduate Student and Research Assistant, *Box 5457, State College Station, Raleigh, North Carolina*

Davis, Richard L., B.S. (North Carolina State) Sales Engineer, *Box 304, Charlotte, North Carolina*

Dickman, Sidney, A.B. (Brooklyn College) Graduate Student at Columbia University, *2823 West 25th Street, Brooklyn 24, New York*

Fitzgerald, Rev. John F., S.J., M.S. (Univ. of Detroit) Assistant Professor of Physics and Mathematics, College of the Holy Cross, Worcester 3, Massachusetts.

Godsey, Ellis B., B.S. (Indiana Univ.) Analytical Statistician, Army Chemical Corps, *1716 Pin Oak Road, Baltimore 4, Maryland*

Ghurye, S. G. M.Sc. (Univ. of Bombay) Student and assistant, Department of Mathematical Statistics, *c/o The Institute of Statistics, Phillips Hall, Chapel Hill, North Carolina*

Gutt, Paul, M.S. (Univ. of Chicago) Ordnance Research #1, Mathematician, *6421 S. Ellis, Chicago, Illinois*

Harman, Harry H., S.M. (Univ. of Chicago) Chief, Statistical Research and Analysis Unit, Personnel Research Section, AGO, Dept. of the Army, *4111 Maryland Ave. (Brookmont), Washington 16, D. C.*

Henderson, Charles R., Ph.D. (Iowa State) Associate Professor, Animal Husbandry Department, Cornell University, Ithaca, New York.

Harter, Harman Leon, Ph.D. (Purdue Univ.) Assistant Professor of Mathematics, Michigan State College, East Lansing, Michigan.

Hoffman, William Charles, M.A. (Univ. of Calif. at Los Angeles) Graduate Assistant, Department of Mathematics, Cornell University, Ithaca, New York.

Hydeman, William Robert, M.A. (Syracuse Univ.) Mathematician, U. S. Navy Department, *3810—39th Street, N.W., Washington 16, D. C.*

Kellerer, Hans, Ph.D. Referent, Bayerisches Statistisches Landesamt, Munchen 8, Rosenheimerstr 130, Germany.

Kramer, Kenneth H., M.S. (Carnegie Inst. of Tech.) Teaching Assistant at Carnegie Institute of Technology, *279 Seneca Street, Turtle Creek, Pennsylvania*

Lieberman, Gerald J., M.A. (Columbia Univ.) Engineer and Mathematical Statistician, Statistical Engineering Laboratory, National Bureau of Standards, Washington 25, D. C.

Lindley, Dennis V., M.A. (Cantab) University Demonstrator in Mathematics, Statistical Laboratory, St. Andrews Hill, Cambridge, England.

Malan, A. P., M.Sc. (South Africa) Professor, U.C.O.F.S., Bloemfontein, South Africa.

Rasch, G., Ph.D. (Copenhagen) Chief of Statistical Department, State Serum Institute, Copenhagen, Denmark

Recao, Manuel Felipe, B.A. (Univ. Venezuela) Director General de Estadistica, Minesterio de Fomento, Professor of Mathematics, Facultad Ciencias Economics, Central University, Calle Real Chacao, Quinta "La Paz," Chacao, Estado Miranda, Venezuela.

Riggs, Charles L., Ph.D. (Univ. of Kentucky) Assistant Professor of Mathematics, Department of Mathematics, Kent State University, Kent, Ohio.

Saxer, Walter, Ph.D. Professor a.d. Eldg. Techn. Hochschule, **Zurich, Goldbach-Kusnacht,** Switzerland.

Scobert, Whitney, M.S. (Univ. of Oregon) Associate Professor of Mathematics, Mathematics Department, Idaho State College, Pocatello, Idaho.

Serfling, Robert E., Ph.D. (Univ. of Mich.) Senior Scientist, Officer in Charge, Statistical Branch, Epidemiology Division, Communicable Disease Center, U. S. Public Health Service, Atlanta, Georgia.

Steyn, Hendrik S., Ph.D. (Univ. of Edinburgh) Lecturer in Statistics, University of Pretoria, *305 Fourth Private Avenue, Villieria, Pretoria, South Africa.*

Zacharias, William B., A.M. (Univ. of Pennsylvania) Instructor in Mathematics, Temple University, *1529—67th Avenue, Philadelphia 26, Pennsylvania*

Zeigler, R. K., Ph.D. (Univ. of Iowa) Associate Professor of Mathematics, Mathematics Department, Bradley University, Peoria 5, Illinois.

REPORT OF THE NEW YORK MEETING OF THE INSTITUTE

The twelfth Annual Meeting of the Institute of Mathematical Statistics was held in New York City on December 27–30, 1949. Headquarters were at the Biltmore Hotel where most of the sessions were held; one or more of the sessions were held at the Hotel Commodore, the McAlpin Hotel, and the Governor Clinton Hotel. The meeting was held in conjunction with the Annual Meeting of the American Statistical Association, the American Association for the Advancement of Science, the American Mathematical Society, the Econometric Society, the Psychometric Society, the Mathematical Association of America, the Association for Computing Machinery, and the American Psychological Association. The following 214 members of the Institute attended:

F. S. Acton, P. H. Anderson, R. L. Anderson, T. W. Anderson, H. E. Arnold, K. J. Arnold Max Astrachan, R. R. Bahadur, E. W. Bailey, T. A. Bancroft, W. D. Baten, E. E. Blanche, C. I. Bliss, R. C. Bose, A. H. Bowker, R. A. Bradley, Dorothy Brady, A. E. Brandt, I. D. J. Bross, T. H. Brown, O. P. Bruno, P. T. Bruyere, R. W. Burgess, J. M. Cameron, B. H. Camp, E. W. Cannon, S. D. Canter, Bernard Carol, O. S. Carpenter, Maria Castellani, Jack Chassan, Randolph Church, Edmund Churchill, W. G. Cochran, A. C. Cohen, Jr., R. H. Cole, E. P. Coleman, F. G. Cornell, Jerome Cornfield, C. C. Craig, M. T. Crapsey, J. F. Daly, D. A. Darling, Besse B. Day, F. R. Del Priore, W. E. Deming, Philip Desind, W. J. Dixon, C. W. Dunnett, Solomon Dutka, P. S. Dwyer, Benjamin Epstein, W. D. Evans, W. T. Federer, William Feller, J. W. Fertig, Leon Festinger, C. H. Fischer, J. C. Flanagan, M. M. Flood, L. R. Frankel, N. M. Franklin, H. A. Freeman, Bernard Friedman, Melitta L. Garbuny, E. F. Gardner, M. A. Geisler, H. H. Germond, Leon Gilford, Abraham Golub, William Gomberg, C. H. Graves, S. W. Greenhouse, J. A. Greenwood, Evelyn S. Grossman, H. T. Guard, Carl Hammer, E. C. Hammond, H. H. Harman, T. E. Harris, Boyd Harshbarger, H. L. Harter, W. A. Hendricks, L. H. Herbach, J. L. Hodges, Jr., Wassily Hoeffding, Helen M. Humes, Harold Hotelling, Cuthbert Hurd, H. M. Hughes, W. R. Hydeman, S. M. Ikhtiar-ul-Mulk, S. I. Isaacson, Marcus Jacobs, W. W. Jacobs, J. E. Jackson, Carol M. Jaeger, J. B. Jeming, R. J. Jessen, H. L. Jones, Alice S. Kaitz, W. C. Kalinowski, Leo Katz, R. D. Keeney, B. F. Kimball, Leslie Kish, Lila F. Knudsen, Paul Koditschek, C. F. Kossack, K. H. Kramer, R. R. Kuebler, Jr., S. M. Kwerel, R. B. Ladd, Marguerite Lehr, F. C. Leone, Joseph Lev, Howard Levene, G. J. Lieberman, Julius Lieblein, S. B. Littauer, Simon Lopata, Irving Lorge, E. D. Lowry, L. H. Madow, W. G. Madow, Benjamin Malzberg, Joseph Mandelson, E. S. Marks, Margaret P. Martin, J. W. Mauchly, P. J. McCarthy, Margaret Merrell,

Albert Mindlin, P. D. Minton, Robert Mirsky, A. M. Mood, Doris N. Morris, R. H. Morris, Dorothy J. Morrow, J. W. Morse, J. E. Morton, Judith Moss, R. G. Moss, Frederick Mosteller, C. M. Mottley, Hugo Muench, L. F. Nanni, Doris Newman, G. E. Noether, M. L. Norden, J. A. Norton, Jr., H. W. Norton, E. G. Olds, P. S. Olmstead, A. L. O'Toole, W. R. Pabst, Jr., R. E. Patton, Katherine Pease, G. W. Petrie, B. E. Phillips, E. W. Pike, Aditya Prakash, Frank Proschan, J. E. Raup, L. J. Reed, J. S. Rhodes, P. R. Rider, H. G. Romig, Norman Rudy, Marion M. Sandomire, F. E. Satterthwaite, Mary Ann Savas, M. A. Schneiderman, Samuel Schweid, O. A. Shaw, G. D. Shellard W. A. Shewhart, S. S. Shrikhande, Harry Shulman, I. H. Siegel, Rosedith Sitgreaves, G. W. Snedecor, Herbert Solomon, D. E. South, Mortimer Spiegelman, R. G. D. Steel, J. R. Steen, Arthur Stein, Joseph Steinberg, F. F. Stephan, A. I. Sternhell, J. S. Stock, J. G. Strieby, J. V. Sturtevant, W. R. Thompson, L. J. Tick, Gerhard Tintner, M. M. Torrey, J. W. Tukey. G. W. Tyler, S. A. Tyler, Uttam Chand, D. F. Votaw, Jr., Helen M. Walker, W. A. Wallis, Samuel Weiss, E. L. Welker, D. R. Whitney, Frank Wilcoxon, R. I. Wilkinson, S. S. Wilks, C. P. Winsor, M. A. Woodbury, Holbrook Working.

The opening session on Tuesday, December 27, 9 A. M., held jointly with the American Statistical Association and the American Mathematical Society, was devoted to *Operations Research*, with Professor J. Steinhardt, Operations Evaluation Group, Massachusetts Institute of Technology presiding. The following papers were presented:

1. *Topics on the Methodology of Operations Research.* B. O. Koopman, Columbia University.
2. *Some Applications of the Mathematical Theory of Games.* G. E. Kimball, Columbia University.
3. *Theory of Games.* L. Gillman, Operations Evaluation Group, Massachusetts Institute of Technology.
4. *Development of Theories of Action.* Ellis Johnson, Operations Research Office. The Johns Hopkins University.
5. *Some Industrial Applications of Operations Research.* A. A. Brown, Operations Evaluation Group, Massachusetts Institute of Technology.

At the second session, held jointly with the American Statistical Association, at 2:30 P. M. on the opening day, Professor M. Loeve, University of California, gave a special invited address entitled, *Fundamental Limit Theorems in Probability*. The discussion was presented by Professor Will Feller of Cornell University and Professor H. E. Robbins of the University of North Carolina. Professor Abraham Wald of Columbia University served as chairman.

The first contributed papers session was held on the same day at 4:00 P. M., with Professor W. D. Baten of Michigan State College and Michigan Agricultural Experiment Station as chairman. The following papers were presented:

1. *The Asymptotic Distribution of the Extremal Quotient.* E. J. Gumbel, New York, and R. D. Keeney, Metropolitan Life Insurance Company, New York.
2. *A Second Formula for Partial Sums of Hyper-geometric Series Having the Unit as Fourth Moment.* Hermann von Schelling, Naval Medical Research Laboratory, New London, Connecticut.
3. *A Coverage Distribution.* Herbert Solomon, Office of Naval Research, Washington, D. C.
4. *The Problem of the Greater Mean.* R. R. Bahadur and Herbert Robbins, University of North Carolina.

5. *Some Extensions of Bayes' Theorem.* F. C. Leone, Case Institute of Technology.
6. *On Optimum Selections from Multinormal Populations.* Z. W. Birnbaum and D. G. Chapman, University of Washington.

On Wednesday morning, December 28, at 10:00 A. M. a session on *Cybernetics* was held jointly with the American Statistical Association and the American Mathematical Society. The following papers were given:

1. *Technique of Multiple Prediction.* Norbert Wiener, Massachusetts Institute of Technology.
2. *Stochastic Problems in Neurophysiology.* Walter Pitts, Massachusetts Institute of Technology.
3. *Information Theory.* Claude Shannon, Bell Telephone Laboratories.

with discussion by Professor J. L. Doob, University of Illinois, Professor Mark Kac, Cornell University, and Professor L. J. Savage, University of Chicago. Professor Jerzy Neyman, University of California was Chairman of the session.

The session on *Review of Statistical Methodology* was held jointly with the American Statistical Association at 2:00 P. M., Wednesday, December 28, with Professor W. A. Wallis, University of Chicago, as chairman. The two papers presented were: *Review of Statistical Methodology in Agriculture and Related Fields*, by Professor W. T. Federer, Cornell University and *Recent Developments in Statistical Methodology in Social Science*, by Professor Frederick Mosteller, Harvard University; discussion followed by Professor L. J. Savage of the University of Chicago.

The second session of contributed papers was held jointly with the American Statistical Association and the Econometric Society on Thursday, December 29, at 10:00 A. M., with Professor H. T. Davis of Northwestern University presiding. The following papers were presented:

1. *Simple Regression Analysis with Autocorrelated Disturbances.* Howard Jones, Illinois Bell Telephone Company.
2. *Application of Sequential Sampling Method to Check the Accuracy of a Perpetual Inventory Record.* Joseph Jeming, New York City.
3. *A Test of Klein's Model III for Changes of Structure.* Andrew Marshall, Rand Corporation.
4. *Application of the Theory of Extreme Values to Economic Problems.* S. B. Littauer, Columbia University and E. J. Gumbel, New York City.
5. *Bias Due to the Omission of Independent Variables in Ordinary Multiple Regression Analysis.* T. A. Bancroft, Iowa State College.
6. *Estimating Parameters of Pearson Type III Populations from Truncated Samples.* A. C. Cohen, Jr., University of Georgia.
7. *The Circular Normal Distribution.* E. J. Gumbel, New York City.

The third session of contributed papers was held at 2:00 P. M. on Thursday, December 29, with Professor L. C. Aroian of Hunter College as Chairman. The following papers were presented in person or by title as indicated:

1. *Treatment of Attenuation Problems by Random Sampling.* H. Kahn and T. Harris, The Rand Corporation.
2. *On the Exstence of Nearly Locally Best Unbiased Estimates.* Herman Rubin, Stanford University.

3. *The Experimental Evaluation of Multiple Definite Integrals.* George Tyler, Naval Electronics Laboratory, San Diego, California.
4. *Tests of Fit of a Cumulative Distribution Function Over Partial Range of Sample Data.* Bradford Kimball, New York State Department of Public Service, New York City.
5. *Large Sample Tests for Comparing Percentage Points of Two Arbitrary Continuous Populations.* A. W. Marshall and John Walsh, The Rand Corporation.
6. *On the Distribution of Wald's Classification Statistics.* Harman L. Harter, Michigan State College.
7. *Analysis of Extreme Values.* W. J. Dixon, University of Oregon.
8. *A Note on the Variance of Truncated Normal Distributions.* (By title) A. C. Cohen, Jr., University of Georgia.
9. *Some Comments on the Efficiency of Significance Tests.* (By title) John Walsh, The Rand Corporation.
10. *Some Estimates and Tests Based on the Smallest Values in a Sample.* (By title) John Walsh, The Rand Corporation.

The subject of the next session, 4:00 P. M. Thursday, December 29, was the *Review of Stochastic Processes from the Point of View of Mathematical Statistics.* This session was held jointly with the American Statistical Association, Professor C. C. Craig of the University of Michigan presiding. Two papers were given, one by Professor A. B. Mann of the National Bureau of Standards, Ohio State University and the University of California; and the second by Professor John Tukey, Princeton University.

On Friday, December 30, at 9:00 A. M. a session on *Statistical Methods in Astronomy* was held jointly with the American Statistical Association and Section D of the American Association for the Advancement of Socience. Professor Walter Bartky of the University of Chicago, Chairman of the session, opened the meeting with introductory remarks on *Astronomical Problems Requiring Statistical Methods.* The following papers were presented:

1. *The Nearby Stars.* Peter Van De Kamp, Swarthmore College.
2. *Corrections to Observed Frequency Distributions.* Bart J. Bok and J. K. De Jonge, Harvard University.
3. *The Problem of Selective Identifiability of Binaries.* Elizabeth Scott, University of California.
4. *Multivariate Periodogram Analysis and Detection of Variable Stars.* Harold Hotelling, University of North Carolina.

These papers were discussed by Professor Jerzy Neyman, University of California.

The session on *Discriminant Functions in Education* was held jointly with the American Statistical Association, the American Psychological Association and the Psychometric Society. Professor T. W. Anderson of Columbia University gave an invited address on *Classification by Multivariate Measures,* followed by discussion by Professors J. C. Flanagan of the University of Pittsburgh and John Carroll of Harvard University. Professor Robert Thorndike of Columbia University presided.

The final session of the meeting was devoted to *Computation* and was held jointly with the American Statistical Association and the Association for Computing Machinery. Professor Harold Hotelling of the University of North Carolina serving as Chairman. The following papers were given:

1. *Idiosyncrasies of Automatically-sequenced Digital Computing Machines.* Ida Rhodes, National Bureau of Standards.
2. *Problem Solving on Large-Scale Automatic Calculating Machines.* W. D. Woo, Harvard University.
3. *A Statistical Application of the UNIVAC.* John Mauchly, Eckert-Mauchly Computer Corporation.

These papers were discussed by James McPherson, Bureau of the Census and Emil Schell, Office of the Air Comptroller.

Meetings of the Council were held on Tuesday, December 27, at 12:00 Noon, Professor Jerzy Neyman presiding and again on Thursday, December 29, at 12:00 Noon, Professor J. L. Doob presiding. The Business Meeting was held on Wednesday, December 28, Professor Jerzy Neyman presiding. The report of this meeting is given elsewhere in this issue.

S. B. LITTAUER,
Associate Secretary

MINUTES OF THE ANNUAL MEMBERSHIP MEETING, NEW YORK, DECEMBER 28, 1949

The meeting was called to order at 4:30 P.M. by President Jerzy Neyman. The annual reports of the President, Editor, and Secretary-Treasurer were read. They are printed elsewhere in this issue.

It was moved by Harold Hotelling that the front cover of the *Annals* in the future shall bear the additional notation that it was edited during the years 1938–1949 by S. S. Wilks. Motion was seconded and carried unanimously.

The tellers reported the election of the following officers:

President-Elect	P. S. Dwyer
Members of the Council for 1950–1952	David Blackwell
	W. G. Madow
	Frederick Mosteller
	L. J. Savage

Meeting was adjourned at 5:15 P.M.

CARL H. FISCHER
Secretary

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1949.

I wish to begin my Report by welcoming the newly elected Fellows, Doctors Z. W. Birnbaum, D. J. Finney, H. O. Hartley, Wassily Hoeffding, Michel Loève, Edward Paulson and S. N. Roy. In addition, a hearty welcome is due to Dr. G. W. Brown who was elected last year, but inadvertently omitted in the published list. The election to the fellowship is a mark of recognition on the part of the Institute. At the same time, I am sure the Institute has reason to be proud of having among its fellows such distinguished scholars as are now added to the list.

During the past year the intensity of the Institute's life grew markedly in many respects. In particular, a very considerable number of our members took part in various Committees. For the sake of brevity, the composition of all the Committees is given in a tabular form at the end of the Report. At this time I wish to express the indebtedness of the Institute to the Chairmen and to the Members of all the Committees.

Undoubtedly the most important function of the members of the Institute is research and the most important function of the Institute itself is the publication of the results of this research. In this respect the past year brought about a fundamental change: after a dozen years of hard and most fruitful work as Editor of the *Annals*, Professor S. S. Wilks resigned this year and the Council elected Professor T. W. Anderson as his successor. According to our present Constitution, the term of office of the Editor is three years.

About a decade ago I suggested and the Membership Meeting of the Institute approved that the cover of the *Annals* bear the name of its founder, Professor Harry Carver. Founded by Carver, the *Annals* were developed by Wilks and, now stand as the most important statistical journal in the world. Accordingly the Chair will welcome a motion to add Professor Wilks' name as a permanent feature of the cover of the *Annals of Mathematical Statistics*.

While being grateful to Wilks and regretting his withdrawal, we should extend a most hearty welcome to T. W. Anderson. Because of his scholarship, broad vision combined with broadmindedness and because of his energy, he is an excellent promise for the future of the *Annals*. It is a pleasure to express the gratitude of the Institute to Columbia University and, in particular, to Dr. Abraham Wald for providing the necessary facilities for the Editorial office of the *Annals*.

Prior to embarking on the election of the new Editor, the I.M.S. Council approved an important document prepared by a special Committee chaired by S. S. Wilks, formulating the editorial policy of the Institute.

Of the many fundamental parts of this document I wish to mention the following:

(i) "In establishing the editorial procedure, special care should be taken to avoid the danger of the Annals becoming a one-group journal rather than serving the Institute as a whole . . . the refusal to publish a paper on grounds of general policy (rather than because of some verifiable defects such as mistakes, triviality, lack of new material, etc.) shall be based on a unanimous agreement of the Editor and of all the Associate Editors."

The general idea behind these passages is, of course, that thus far, the *Annals* is the only journal published by the Institute and should provide facilities for all the different schools of thought. My understanding is that this includes the biostatistician Cochran and the econo-statistician Koopmans, the multivariate Hotelling and the tolerant Wilks, the quality-control-minded Shewhart and the dependently-limiting Loève, the necessary- and sufficient-normal Feller and the minimax-gambler von Neumann, the relativistically-cybernetic Wiener and the general-sequential-decision-maker Wald. I should think that even our next

President, the stochastically-processed-Markovian Doob, is meant to have a chance to publish in the *Annals*, from time to time.

(ii) Another interesting point in the same document concerns the proposed approximate distribution of space in the Annals:
- (a) research papers on mathematical statistics proper—60 per cent;
- (b) research papers in borderline fields, including applications—20 per cent;
- (c) expository papers—15 per cent
- (d) news, notices, etc.—5 per cent.

Since in the past there was too little expository material, the Council instituted the so-called Special Inivited Papers, to be presented from time to time on selected subjects. The text of these papers, accompanied by the prepared discussion, will be printed in the *Annals*. The program of the present meeting includes our first Special Invited Paper, by Michel Loève. It is hoped that the Special Invited Papers will satisfy the need for expository material now felt by the membership of the Institute. I am sure the Program Committees will appreciate suggestions of the Members regarding the sections of the theory requiring expository presentation.

The financial aspect of the publication program of the Institute was a continued worry of the Council. As is well known, the *Annals* is overloaded with papers and the cost of printing is growing constantly. In order to ease the situation somewhat, our new Constitution was amended to include the provision that the Universities and other institutions could become Institutional Members. There is already some additional income from this source and, if all the members of the Institute are energetic in urging their Departments to become Institutional Members, this income may be quite substantial.

It is conceivable that some potential sources of funds exist, not directly available for the *Annals*, which may be used for starting a new statistical journal. In order to investigate this possibility a special committee was appointed under the chairmanship of Professor Scheffé. This Committee did an excellent job in trying to find a solution of the tremendously difficult problem and there is now a reasonable hope that, in the not very distant future, our publication facilities will be increased.

Another deep change in the structure of the Institute occurred this year. Here I have in mind the resignation of Dr. Paul S. Dwyer, our long and hard working Secretary, and the taking over by Dr. Carl Fischer. Dr. Dwyer's resignation was announced last year at the meeting at Cleveland and we expressed to him our hearty thanks for his untiring work for the Institute. I wish to repeat these thanks now and to accompany them by the hearty congratulations on the excellent program he prepared for this meeting in his new capacity as the Chairman of the National Program Committee.

Until recently, there was a certain disequilibrium in the location of the meetings of the Institute. Practically all of the meetings were held in the East and the West Coast members could attend them only as a matter of exceptional luck. Later, regional meetings were organized, and this year we have functioning three

Regional Program Committees, one for the East, one for the West Coast and one for the Middle West. In addition, we have Program Committees for the two National Meetings of the Institute. In parallel with the redistribution of meetings, there was an increase in their number. This process was accompanied by the very efficient help on the part of the governmental organizations, of the Office of Naval Research, the Air Force, and the Army, for the members of the Institute to attend the meetings even if they are held at a considerable distance. As a combined result of these developments it now may seem that there are too many meetings. Undoubtedly, the number and the location of future meetings of the Institute will be seriously discussed and adjusted to the existing needs.

Naturally, the help of the Governmental institutions was not limited to help in travel. A considerable number of research projects in statistics are now in progress in many institutions with excellent results for science, for the younger people who are given the chance to make their first independent research work without undue worry about food and shelter and, thus, for the country as a whole. The first organization to support fundamental research in general, and in statistics, in particular, seems to be the office of Naval Research. Its broadmindedness and understanding of the spirit of research have established a very high standard which is also sustained by other institutions. If permitted to function as they do now, these institutions will mark an epoch in the development of scholarly work in this country.

The following persons have accepted the appointment to the Nominating Committee for the next year

Henry Scheffé—*Chairman*
Albert W. Bowker
Paul G. Hoel
Leonid Hurwicz
Herbert E. Robbins
David F. Votaw, Jr.

Composition of the Committees of the Institute in 1949

1. *Program Committees* (P.C.)

(i) Eastern P.C. for the April 1949 meeting in New York
Churchill Eisenhart, *Chairman*
W. G. Cochran
C. F. Kossack
S. B. Littauer
F. Mosteller

(ii) West Coast P. C. for June meeting in Berkeley
M. A. Girshick, *Chairman*
Z. W. Birnbaum
W. J. Dixon
J. L. Hodges, Jr.
P. G. Hoel
A. M. Mood

(iii) National P.C. for the Summer Meeting at Boulder, Colorado
W. Feller, *Chairman*

(iv) Mid West P.C.
C. C. Craig. *Chairman*

J. L. Doob
M. A. Girshick
C. C. Hurd
J. Wolfowitz
L. Hurwicz
W. G. Madow
K. May
L. J. Savage
D. R. Whitney

(v) National P.C. for the December meeting in New York
P. S. Dwyer, *Chairman*
J. Berkson
G. W. Brown
C. Eisenhart
Mark Kac
H. Rubin

(vi) Eastern P.C. for the Spring 1950 meeting in North Carolina.
H. Hotelling, *Chairman*
D. Blackwell
H. Geiringer
S. B. Littauer
D. F. Votaw, Jr.
S. S. Wilks

2. *Committee for Special Invited Papers*
J. W. Tukey, *Program Coordinator, Chairman ex officio*
C. C. Craig
P. S. Dwyer
C. Eisenhart
W. Feller
M. A. Girshick
H. Hotelling

3. *Committee on Editorial Policy (1948–1949)*
S. S. Wilks, *Chairman*
W. G. Cochran
W. Feller
M. A. Girshick
P. S. Olmstead
J. Neyman
W. A. Wallis
J. Wolfowitz

4. *Committee to Nominate Candidates for the Editor of the Annals*
Harry C. Carver, *Chairman*
David Blackwell
S. Lee Crump
Erich L. Lehmann
Howard Levene
Frederick Mosteller
Herbert E. Robbins

5. *Committee on Tabulation*
C. Eisenhart, *Chairman*
C. I. Bliss
F. W. Dresch
H. H. Germond
H. O. Hartley
C. C. Hurd
A. N. Lowan
W. G. Madow
H. G. Romig
L. E. Simon

6. *Committee on Directory*
John W. Tukey, *Chairman*
Churchill Eisenhart

7. *Committee to Revive the Statistical Research Memoirs*
Henry Scheffé, *Chairman*
T. W. Anderson
Walter Bartky
C. C. Hurd
George Kuznets

8. *Rietz Lectures Committee*

The Chairmanship of this Committee was accepted by Abraham Wald, the first Rietz Lecturer, who undertook to make further appointments. These are:

C. C. Craig
W. Feller

9. *Committee to Encourage Membership outside of the United States*

T. W. Anderson, *Chairman*
C. C. Hurd
M. Loève
J. Marschak

10. *Committee on Statisticians in the Government Service*

W. E. Deming, *Chairman*
C. Eisenhart

11. *Representative of the I.M.S. to the American Association for the Advancement of Science*

Harold Hotelling

12. *Representative of the I.M.S. to the National Research Council, Division of Physical Sciences*

Walter Bartky (1948–1950)

13. *Representative of the I.M.S. to the Mathematical Policy Committee*

S. S. Wilks

14. *Representative of the I.M.S. to the Joint Committee for Development of Statistical Applications in Engineering and Manufacturing*

Benjamin Epstein

15. *Representatives to the Inter-Society Cooperation on Mathematical Training of Social Scientists*

T. W. Anderson
J. L. Doob
S. S. Wilks

16. *Committee to Determine the Duties and Responsibilities of the Program Committees*

Harold Hotelling, *Chairman*
M. A. Girshick
S. B. Littauer

J. Neyman
President

December 31, 1949

REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE FOR 1949

At the beginning of 1949 the Institute had 1101 members and during the period covered by this report 153 new members (8 of whom begin their membership with 1950) joined the Institute and two members were re-instated. During 1949 the Institute lost 87 members of which 27 were by resignation and 60 by

suspension for non-payment of dues. Judging from the information available at this date, the Institute will have 1167 members as it starts 1950.

During 1949 the Constitution was amended to provide for a new class of membership: Institutional Membership. Although the campaign for institutional members started late in the year, by December 31 there were five universities on the rolls: California, Purdue, Illinois, Princeton and North Carolina. It is hoped that many more universities and corporations will enroll during 1950.

Meetings of the Institute held during 1949 included those at Columbia University on April 8–9, at the Berkeley campus of the University of California on June 16–18, at the University of Colorado on August 29–September 1, and at New York City on December 27–30. The Secretary wishes to call attention to the excellent work of the members who served as Assistant and Associate Secretaries at these meetings: Professor S. B. Littauer at New York, Professor J. L. Hodges, Jr., at California, Professor H. T. Guard at Colorado and Associate Secretary Professor Littauer who was responsible for the New York Meeting.

The following Fellows served as members of the Committee on Fellows: C. C. Craig, chairman, T. W. Anderson, M. A. Girshick, Harold Hotelling, Henry Scheffé, and F. F. Stephan.

The meeting scheduled for November 25–26 at the University of California at Los Angeles was cancelled by vote of the West Coast membership because of the proximity of the Boulder and Christmas Meetings.

At the Council meeting at Boulder, August 29, 1949, the following Associate Secretaries were elected:

Associate Secretary	*Section*
S. B. Littauer	Eastern
K. J. Arnold	Central
J. L. Hodges, Jr.	Western

By a mail vote of the Council, conducted during October, 1949, T. W. Anderson was elected Editor for the period 1950–1952.

A summary of the financial status of the Institute is given below:

FINANCIAL STATEMENT

December 20, 1948 to December 31, 1949

A. RECEIPTS

Balance on Hand,* December 20, 1948	$ 7,121.01
Dues	7,826.35
Contributions	156.15
Life Memberships	392.50
Institutional Memberships	400.00
Subscriptions	4,779.07
Sale of Back Issues	3,314.41
Biometrika	793.50
Income from Investments	100.00
Miscellaneous	169.70
Total	$25,052.69

* In bank deposits and government bonds.

B. EXPENDITURES

Annals—Current		
Office of the Editor	$ 275.00	
Waverly Press	8,777.65	$ 9,052.65
Annals—Back Numbers		
Reprinted Vol. II #4; III #4; IV #3 & #4; V #1; VI #1, 2, 3 & 4; XIII #1, 2, & 4		$ 2,910.55
Mathematical Reviews and Inter-Society Committee		206.92
Office of the Secretary-Treasurer		
Printing, memoranda, etc. (Including some stamped envelopes)	$1,156.61	
Postage, supplies, express, telephone calls	275.00	
Clerical help	2,208.40	
Travelling expense	223,61	$ 3,863.62
Miscellaneous		$ 379.57
Biometrika		$ 657.30
Balance on Hand, *December 31, 1949		$ 7,982.08
Total		$25,052.69

C. SUMMARY OF RECEIPTS AND EXPENDITURES

Balance on Hand, *December 20, 1948	$ 7,121.01
Receipts during 1949	17,931.68
Expenditures during 1949	17,070.61
Balance on Hand, *December 31, 1949	$ 7,982.08

D. LIFE MEMBERSHIP FUNDS

It has been the practice to set up an amount equal to all life membership payments as a liability and to hold all these funds in reserve until the death of the member—after which his payment is released to the general fund. There were three new life membership payments in 1949.

	December 20, 1948	*December 31, 1949*
Number of Life Members	29	32
Total Reserve Held	$2,280.00	$2,672.50

E. BACK ISSUES FUND

It has been our policy, since January 1, 1948, to use income from the sale of back issues to finance the additional reprinting of back issues.

Previous balance in back issues fund	$ 749.77
Income from the sale of back issues during 1949	3,314.41
Expense for reprinting back issues in 1949	2,910.55
Balance, December 31, 1949	$1,153.63

F. BALANCE SHEET, DECEMBER 31, 1949

ASSETS

	December 31, 1949	*Increase since December 20, 1948*
Cash	$ 3,094.08	$ 861.07
U. S. Government G Bonds	3,000.00	—
U. S. Government F Bonds (Purchase price)	1,888.00	—
Current Accounts Receivable	545.78	254.56
Estimated Value (Cost of Back Annals**)	16,459.22	3,673.61
	$24,987.08	$4,789.24

* In bank deposits and government bonds

** Cost of *Annals* calculated at 67 cents per copy

LIABILITIES

Reserve for Life Memberships	$ 2,672.50	$ 392.50
Reserve for Reprinting Back Issues	1,153.63	403.86
Surplus	21,160.95	3,992.88
	$24,987.08	$4,789.24

G. SUMMARY

The surplus of the Institute has increased during the year of 1949 by $3,992.88. While this indicates a favorable condition, it should be noted that roughly 92% of this gain is represented by an increase in the inventory of back issues of the Annals. This asset is definitely of the non-liquid sort and thus the major portion of our gain is of little assistance in meeting our current need for more publication space in the *Annals*.

It should be noted that the year-end statements have always included a substantial amount in prepaid dues and subscriptions on the asset side without a corresponding liability. The figure for December 20, 1948 is $4,060.50 and for December 31, 1949 is $4,682.37. Thus it will be seen that we are virtually running on a hand-to-mouth basis. It is hoped that an increase in the number of individual and institutional memberships during 1950 will bring us into a more favorable situation.

Beginning with January 1, 1950 we plan to revise the bookkeeping system which is no longer adequate for an organization of our present size. In the future, these reports will be made on an accural basis rather than a cash basis and thus will present the data pertaining to each year on a more realistic basis.

We are now in a position to supply all issues beginning with Volume 1. Five or six of the back issues are in short supply, but we expect to be able to reprint these when our supplies become exhausted, using receipts from the sale of back issues to pay for the reprinting.

CARL H. FISCHER
Secretary-Treasurer

December 31, 1949

REPORT OF THE EDITOR OF THE ANNALS FOR 1949

The 1949 volume of the *Annals* exceeded, by a few pages, the 600 pages budgeted for it at the beginning of the year. A total of 65 papers were published, as well as the usual reports, abstracts, and items of news and notices. The 1949 volume was Volume 20 of the *Annals*, and it seemed fitting to publish a cumulative index of papers for the first twenty volumes of the *Annals*. Such an index, containing both author and subject indexes, has been published as a separate 31-page pamphlet and is being distributed with the December 1949 issue of the *Annals*.

The rate of submission of manuscripts continues to increase. By the end of 1949 enough manuscripts to fill two issues of the *Annals* had been accepted for publication. At the same time approximately forty manuscripts were at various stages of refereeing and revision. This means that authors submitting manuscripts at the beginning of 1950 can hardly expect to see their papers in print in less than a year. The rate at which the average gap between submission of manuscripts and their appearance in print has, for the last two years, increased about two issues (six months) per year. There is no reason to predict that this rate will change for at least another year or two. Thus, it is highly desirable that every effort be made to expand the publication program of the Institute during 1950.

The most immediate possibility would be to expand the *Annals* by at least 100 pages if the budget will permit. In the meantime, it is hoped that the Institute committee to study the feasibility of reviving the *Statistical Research Memoirs* will be able to work out a practical plan for further increasing the publication facilities of the Institute.

The manuscripts being submitted continue to cover a wide range of topics in probability and statistics. There is still a scarcity of good review and expository articles being submitted, but with the institution of special invited addresses so widely discussed at the Cleveland meeting of the Institute in December, 1948, we can expect to receive more review and expository articles in the future.

The Editor takes this opportunity to acknowledge, on behalf of the Editorial Committee, the refereeing assistance which has been generously given during the year by the following persons: A. C. Aitken, E. W. Barankin, Z. W. Birnbaum, R. C. Bose, A. H. Bowker, G. W. Brown, K. L. Chung, W. J. Dixon, A. Dvoretzsky, Hilda Geiringer, L. A. Goodman. T. N. E. Greville, F. E. Grubbs, John Gurland, M. H. Hansen, T. E. Harris, H. O. Hartley, E. L. Kaplan, B. F. Kimball, T. Koopmans, Julius Lieblein, H. Levene, M. S. MacPhail, P. J. McCarthy, R. B. Murphy, G. E. Noether, E. G. Olds, P. S. Olmstead, Richard Otter, E. Paulson, M. P. Peisakoff, E. J. G. Pitman, Milton Sobel, D. F. Votaw, Max Woodbury, and J. L. Walsh.

Thanks are due to Mr. M. E. Freeman, Mr. L. A. Goodman and Mr. E. F. Whittlesey for preparation of manuscripts and to Mrs. Lily D. Smith for other editorial and office assistance in connection with the *Annals*.

Finally, on behalf of the Editorial Board, which has had the responsibility for editing the *Annals* since 1938, the Editor extends every good wish to the new Editor, T. W. Anderson, and the new Editorial Board, who will inherit nearly a full year of accepted manuscripts but will otherwise assume editorial responsibility for the *Annals* beginning with the 1950 volume.

S. S. Wilks
Editor.

December 21, 1949